

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:39:13 : Search time 11.1 seconds
(without alignments)
788.344 Million cell updates/sec

Title: US-09-441-723-1
Perfect score: 1185
Sequence: 1 MGPLPRTVELFYDLSPYSW.....AHLGKWMGPPIPAVNARL 226

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1180	99.6	225	1	GTK1_HUMAN
2	869	73.3	225	1	GTK1_RAT
3	328.5	27.7	226	1	YS21_GAEEL
4	298.5	25.2	225	1	YS17_GAEEL
5	95	8.0	203	1	NAHD_PSEPU
6	89.5	7.6	962	1	PTRA_ECOLI
7	89	7.5	212	1	NAHD_PSESP
8	89	7.5	592	1	THD1_ARATH
9	88.5	7.5	613	1	SG2_BOVIN
10	84.5	7.1	316	1	MIAA_TREPA
11	84.5	7.1	971	1	S3B1_YEAST
12	83.5	7.0	488	1	CATA_LISSE
13	83.5	7.0	700	1	NCD_DRONE
14	81.5	6.9	619	1	SG2_RAT
15	81.5	6.9	1061	1	DPOL_ADE12
16	81	6.8	274	1	YG73_SYNY3
17	81	6.8	966	1	ST10_MOUSE
18	80.5	6.8	705	1	CC5_YEAST
19	80.5	6.8	1073	1	HSEB_PIG
20	80	6.8	540	1	TOPI_AOUAE
21	80	6.8	1162	1	BXEN_CLOBO
22	79.5	6.7	547	1	CATX_BAGSU
23	79.5	6.7	1073	1	USER_HUMAN
24	78.5	6.6	383	1	PNAH_RICPR
25	78	6.6	1162	1	BXEN_CLOBO
26	77.5	6.5	634	1	YC36_METUA
27	77	6.5	494	1	ADRO_RAT
28	76.5	6.5	962	1	YBX7_SCHPO
29	76.5	6.5	1056	1	DPOL_ADE02
30	76.5	6.5	1193	1	DPOL_ADE04
31	76	6.4	406	1	PKK3_HUMAN
32	76	6.4	646	1	NODQ_RHIS3
33	76	6.4	2208	1	POLN_WANCV

34	75.5	6.4	744	1	GYRA_AOUAE
35	75.5	6.4	887	1	GLND_KLEPN
36	75.5	6.4	1056	1	DPOL_ADE05
37	75.5	6.4	1398	1	PLS_PYRFU
38	75.5	6.4	1434	1	RPOB_UREPA
39	75	6.3	231	1	LOLD_NEIMA
40	75	6.3	393	1	HEMX_ECOLI
41	75	6.3	593	1	KPYA_TOBAC
42	75	6.3	631	1	OE66_NPVLS
43	75	6.3	686	1	MEPD_PIG
44	75	6.3	925	1	PIPL_YEAST
45	75	6.3	984	1	SECA_AOUAE

ALIGNMENTS

RESULT	1	GTK1_HUMAN	STANDARD:	PRT:	225 AA.
ID	Q9Y2Q3	Q9P1S4			
AC	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DE	Glutathione S-transferase, mitochondrial (EC 2.5.1.18) (GST 13-13)				
DE	Glutathione S-transferase subunit 13 (GST class-kappa) (HDCMD47P).				
GN	GTK1				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Blood;				
RX	MEDLINE=20499367; PubMed=11042152;				
RA	Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,				
RA	Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,				
RA	Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.,				
RT	"Cloning and functional analysis of cDNAs with open reading frames for				
RT	300 previously undefined genes expressed in CD34+ hematopoietic				
RT	stem/progenitor cells."				
RT	Genome Res. 10:1546-1560(2000).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	Zhao Z., Huang X., Li N., Zhu X., Cao X.;				
RA	"A novel gene from human dendritic cell."				
RT	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
RL	[3]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Uterus;				
RC	MEDLINE=21134917; PubMed=11230166;				
RX	Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,				
RA	Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,				
RA	Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,				
RA	Mewes H.-W., Ottenwälder B., Obermaier B., Tampe J., Heubner D.,				
RA	Wamburt R., Korn B., Klein M., Poustka A.;				
RT	"Towards a catalog of human genes and proteins: sequencing and				
RT	analysis of 500 novel complete protein coding human cDNAs.";				
RT	Genome Res. 11:422-435(2001).				
CC	-1- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC				
CC	ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT (BY SIMILARITY).				
CC	-1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.				
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).				
CC	-1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY, KAPPA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/				
CC	or send an email to license@isb-sib.ch).				

067108	aquifex aeo
P41393	klebsiella
P04495	human adeno
P72186	pyrococcus
Q98V66	ureaplasma
P57030	neisseria m
P09127	escherichia
Q05455	nicotiana t
P89518	leucania se
P47788	sus scrofa
P40020	saccharomyc
067718	aquifex aeo

CC EMBL; AF070657; AAD20963.1; -
 DR EMBL; AF068287; AAF65506.1; -
 DR EMBL; AL136938; CAB66872.1; -
 DR MIM; 602321; -
 DR InterPro; IPR004287; HCCA_isomerase.
 DR Pfam; PF03046; HCCA_isomerase; 1.
 KW Transferase; Mitochondrion.
 FT INIT_MET 0 BY SIMILARITY.
 FT CONFLICT 178 178 G -> R (IN REF. 2).
 FT CONFLICT 219 219 P -> S (IN REF. 2).
 SQ SEQUENCE 225 AA; 25365 MW; FE91A5EE0F0B0A1 CRC64;

Query Match 99.6%; Score 1180; DB 1; Length 225;
 Best Local Similarity 100.0%; Pred. No. 6.7e-95;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLPR 61
 DB 1 GPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLPR 60
 QY 62 KGLYMANDLKLRHHLQIPHPKDFLSVMEKGSLSAMRFLTAVNLEHPEMLEKASREL 121
 DB 61 KGLYMANDLKLRHHLQIPHPKDFLSVMEKGSLSAMRFLTAVNLEHPEMLEKASREL 120
 QY 122 WMRVWSRNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 181
 DB 121 WMRVWSRNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 180
 QY 182 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 226
 DB 181 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 225

RESULT 2
 ID GTK1_RAT STANDARD; PRT; 225 AA.
 AC P24473; O09034;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutathione S-transferase, mitochondrial (EC 2.5.1.18) (GST 13-13)
 GN GSTK1 OR GSTK1-1.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97079244; PubMed=8920976;
 RA Pembler S.E., Wardle A.F., Taylor J.B.;
 RT "Glutathione S-transferase class kappa: characterization by the
 RT cloning of rat mitochondrial GST and identification of a human
 RT homologue.";
 RL Biochem. J. 319:749-754 (1996).
 RN [2]
 RP SEQUENCE OF 1-33.
 RX MEDLINE=91354194; PubMed=1883325;
 RA Harris M.J., Meyer D.J., Coles B., Ketterer B.;
 RT "A novel glutathione transferase (13-13) isolated from the matrix of
 RT rat liver mitochondria having structural similarity to class theta
 RT enzymes.";
 RL Biochem. J. 278:137-141 (1991).
 CC -!- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC
 CC ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT.
 CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 or send an email to license@isb-sib.ch).
 DR EMBL; S83436; AAB50831.1; -
 DR PIR; S17164; S17164.
 DR InterPro; IPR004287; HCCA_isomerase.
 DR Pfam; PF03046; HCCA_isomerase; 1.
 KW Transferase; Mitochondrion.
 FT INIT_MET 0
 FT CONFLICT 1 1 G -> C (IN REF. 2).
 SQ SEQUENCE 225 AA; 25362 MW; FC895B730655E0C9 CRC64;

Query Match 73.3%; Score 869; DB 1; Length 225;
 Best Local Similarity 69.3%; Pred. No. 4.7e-58;
 Matches 156; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLPR 61
 DB 1 GPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLPR 60
 QY 62 KGLYMANDLKLRHHLQIPHPKDFLSVMEKGSLSAMRFLTAVNLEHPEMLEKASREL 121
 DB 61 KGLYMANDLKLRHHLQIPHPKDFLSVMEKGSLSAMRFLTAVNLEHPEMLEKASREL 120
 QY 122 WMRVWSRNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 181
 DB 121 WMRVWSRNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 180
 QY 182 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 226
 DB 181 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 225

RESULT 3
 ID YS21_CAEEL STANDARD; PRT; 226 AA.
 AC Q09652;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 26.5 kDa protein 2K1320.1 in chromosome II.
 GN 2K1320.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 CC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Berks M.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 or send an email to license@isb-sib.ch).
 DR EMBL; Z46934; CAA87039.1; -
 DR WormPep; ZK1320.1; CE01698.
 DR InterPro; IPR004287; HCCA_isomerase.
 DR Pfam; PF03046; HCCA_isomerase; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 226 AA; 26482 MW; DBA39A5994300164 CRC64;

us-09-441-723-1-rsp

Tue May 21 08:04:02 2002

120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDHLVQDVLMVMSDSAEVKNILRENTDEAIGN 176

179 GAFGLP---ITVAHVDGQT-HMLFGSDRMELLALHLLGKWMGPI 218

177 GCFGAPWMTDGH--GKVLQTVFGSDRLPOVADFLAEFFKGP 218

RESULT 5

NAHD_PSEPU STANDARD; PRT; 203 AA.

AC Q51948; 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE 2-hydroxychromene-2-carboxylate isomerase (HCCA isomerase).

GN NAHD.

OS Pseudomonas putida.

OG Plasmid NAH7.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=G7 / ATCC 17485;

RX MEDLINE=95095951; PubMed=8002605;

RA Eaton R.W.;

RT "Organization and evolution of naphthalene catabolic pathways; sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and trans-o-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid."

RL J. Bacteriol. 176:7757-7762(1994).

RN [2]

RP CHARACTERIZATION.

RC STRAIN=G7 / ATCC 17485;

RX MEDLINE=93077433; PubMed=1447127;

RA Eaton R.W., Chapman P.J.;

RT "Bacterial metabolism of naphthalene: construction and use of recombinant bacteria to study ring cleavage of 1,2-dihydroxynaphthalene and subsequent reactions."

RL J. Bacteriol. 174:7542-7554(1992).

CC -1- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THPPA).

CC THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.

CC -1- COPACTOR: GLUTATHIONE SEEMS TO STABILIZE THE ENZYME, WHICH LOSES ACTIVITY RAPIDLY IN THE ABSENCE OF THIS COMPOUND.

CC -1- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES CONVERSION OF NAPHTHALENE TO SALICYLATE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U09057; AAA66358.1;

DR InterPro: IPR004287; HCCA_isomerase.

DR Pfam: PF03046; HCCA_isomerase; 1.

KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.

SQ SEQUENCE 203 AA; 23061 MW; E84B56F21C604945 CRC64;

Query Match 8.0%; Score 95; DB 1; Length 203;

Best Local Similarity 24.2%; Pred. No. 0.28;

Matches 51; Conservative 35; Mismatches 97; Indels 28; Gaps 8;

QY 8 VELYDVLSPYSWLGFEILCRYNQINWINLQRLSLITGIMKDSGNKPPGL--LPRKGLY 65

DB 3 VDFYDFLSPFYSYLANQRLSKLAQDYGITIRYNAIDARVKIATGNVGSNRDLKVKLDY 62

QY 66 MANDLKLRLHULOIPHFHFKDFLSVMLEK---SLSAMRFTFAVLEHPEMLEKASRELW 122

Query Match 27.7%; Score 328.5; DB 1; Length 226;

Best Local Similarity 33.0%; Pred. No. 2.2e-21;

Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVELFVDVLSYSLGFEILCRYNQINWINLQRLSLITGIMKDSGNKPPGL 59

DB 1 MPKLPRI-IDFYDVLSYSLGFEILCRYNQINWINLQRLSLITGIMKDSGNKPPGL 59

QY 60 PRKGLYMANDLKLRLHULOIPHFHFKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119

DB 60 PARSIMMTDLKRTAKFWIDPPLPFLPFWIWKYRTGAMKVLVLLVLEQDKELMLRAAR 119

QY 120 ELWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEACRYG 179

DB 120 EMWVRLSRSEKIFEDDFEVLKAVGV--KNPEQIVKESKDEKIKITLMNTNKGVDLM 177

QY 180 AFGLP-ITVAHVDGQTHMFGSDRMELLALHLLGKWMGPI 219

DB 178 AYGAWINVHTEDGSEHSFFGSDRFLIADLLQPP--QPLP 216

RESULT 4

YIS7_CABEL STANDARD; PRT; 225 AA.

AC Q18973; 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 26.0 kDa protein D2024.7 in chromosome IV.

GN D2024.7

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX Du 2., Gattung S.;

RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U41011; AAA82289.1;

DR WormPep: D2024.7; CE04296.

DR Pfam: PF03046; HCCA_isomerase; 1.

KW Hypothetical protein.

SQ SEQUENCE 225 AA; 25951 MW; 23D0A3D6762B7232 CRC64;

Query Match 25.2%; Score 298.5; DB 1; Length 225;

Best Local Similarity 33.0%; Pred. No. 8.4e-19;

Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;

QY 3 PLPRTVELFVDVLSYSLGFEILCRYNQINWINLQRLSLITGIMKDSGNKPPGL--L 59

DB 2 PNRRVVRFFFDVLSYSLGFEILCRYNQINWINLQRLSLITGIMKDSGNKPPGL--L 59

QY 60 PRKGLYMANDLKLRLHULOIPHFHFKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119

DB 60 PIKEYMHKDLFLSAQYWGIPRLPKOVNTNMLNTSSIVPQRTILVASQLDRNVLMDVAR 119

QY 120 ELWMRVWSRNEDI--TEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEAACRY 178

DB 63 LKVDLQWQAQYGLPVPANYSRRNIGFYYSAGAAQAAVNV-----VF 110

QY 123 MRVMSRNDITEPOSILAA--AEKAGMSAEQAQGLLEKATPKVKNOLKETTEAACRYGA 180

DB 111 NAVM--CEGIAPDLESPLALYSEKLGWDRSAFEHFLUSSNAATE---RYDEQTHAAIERKV 165

QY 181 FGLPITVAHVDGQTHMLFGSDRMELLAHLG 211

DB 166 FGVP-TMFLGD---EMWGNDRFLMESAMG 192

RESULT 6

ID PTRA_ECOLI STANDARD; PRT; 962 AA.

AC P05458; P78106;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protease III precursor (EC 3.4.24.35) (Pitriylsin) (Protease pi).

GN PTRA OR PTR OR B2821.

OS Escherichia coli.

OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID-562;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE-87040734; PubMed-3534791;

RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emmerson P.T.;

RT "Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease III.";

RL Nucleic Acids Res. 14:7695-7703(1986).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE-97426617; PubMed-9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [3]

RN SEQUENCE OF 853-962 FROM N.A.

RX MEDLINE-87066729; PubMed-3537960;

RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,

RA Emmerson P.T.;

RT "Complete nucleotide sequence of the Escherichia coli recB gene.";

RL Nucleic Acids Res. 14:8573-8582(1986).

RN [4]

RN SEQUENCE OF 1-296 FROM N.A.

RC STRAIN-K12.

RX MEDLINE-88005781; PubMed-3308636;

RA Claverie-Martin F., Diaz-Torres M.R., Kushner S.R.;

RT "Analysis of the regulatory region of the protease III (ptr) gene of Escherichia coli K-12.";

RL Gene 54:185-195(1987).

RN [5]

RN MUTAGENESIS, AND ACTIVE SITE.

RX MEDLINE-92237263; PubMed-1570301;

RA Becker A.B., Roth R.A.;

RT "An unusual active site identified in a family of zinc metalloendopeptidases.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).

RN [6]

RN MUTAGENESIS, AND ACTIVE SITE.

RX MEDLINE-93277449; PubMed-8099278;

RA Becker A.B., Roth R.A.;

RT "Identification of glutamate-169 as the third zinc-binding residue in proteinase III, a member of the family of insulin-degrading enzymes.";

RL Biochem. J. 292:137-142(1993).

CC -!- FUNCTION: ENDOPEPTIDASE THAT DEGRADES SMALL PEPTIDES OF LESS THAN 7 kDa, SUCH AS GLUCAGON AND INSULIN.

CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF 16-TYR-|-LEU-17 AND 25-PHE-|-TYR-26 BONDS OF OXIDIZED INSULIN B CHAIN. ALSO ACTS ON OTHER SUBSTRATES OF MW LESS THAN 7 kDa SUCH AS INSULIN AND GLUCAGON.

CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC.

CC -!- SUBUNIT: MONOMER.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16; ALSO KNOWN AS THE INSULINASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. Usage by and for commercial or send an email to license@lsb-sib.ch).

CC -----

CC EMBL; X04581; CAA28249.1; -

CC EMBL; U29581; AAB40468.1; -

CC EMBL; AE000365; AAC75860.1; -

CC EMBL; X06227; CAA29576.1; -

CC EMBL; M17095; AAA24436.1; -

CC PIR; A29093; SNECFI.

CC MEROPS; M16.001; -

CC EC02DBASE; G095.0; 6TH EDITION.

CC EcoGene; EGI0786; ptrA.

CC InterPro; IPR001431; Peptidase_M16.

CC Pfam; PF00675; Peptidase_M16; 1.

CC PROSITE; PS00143; INSULINASE; 1.

CC KW Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;

CC Complete proteome.

CC SIGNAL 1 23

CC CHAIN 24 962 PROTEASE III.

CC METAL 88 88 ZINC.

CC ACT_SITE 91 91 ZINC.

CC METAL 92 92 ZINC.

CC METAL 169 169 ZINC.

CC MUTAGEN 88 88 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.

CC MUTAGEN 91 91 E->Q: LOSS OF ACTIVITY.

CC MUTAGEN 92 92 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.

CC MUTAGEN 162 162 E->Q: 20% LOSS OF ACTIVITY.

CC MUTAGEN 169 169 E->Q: LOSS OF ACTIVITY AND OF ZN-BINDING.

CC MUTAGEN 204 204 E->Q: NO LOSS OF ACTIVITY.

CC CONFLICT 277 284 IIIHYVPA -> HYHSLRPW (IN REF. 4).

CC SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;

Query Match 7.6%; Score 89.5; DB 1; Length 962;

Best Local Similarity 26.3%; Pred. No. 5.4; Mismatches 48; Indels 23; Gaps 3;

Matches 30; Conservative 13;

QY 126 WSRNDEITEPQSTILAAAEKAGMSAEQA-----QGLEKIATPKVKNQ 167

DB 731 WCRNKDVVVKKQSVIFEKAGNSTDSALAAVFVPTGYDEYTSAYSSLLGIVQPFYNO 790

QY 168 LKETTEAACRYGAFGLPITVAHVDGQTHMLFGSDRMELLAHLGKRWGPIPPA 221

DB 791 LR--TEEQLGVAVFAPFMSVGVGROWMGFLLQSDKQ---PSFLWERYKAFFPTA 839

RESULT 7

NAHD_PSESP STANDARD; PRT; 212 AA.

ID NAHD_PSESP

AC Q52462;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE 2-hydroxychromene-2-carboxylate isomerase (HCCA isomerase).

GN DOXJ

OS Pseudomonas sp. (strain C18).

OG Plasmid.

OC Bacteria; Proteobacteria.

```

NCBI_TaxID=306;
[1]
NCBI_TaxID=306;
SEQUENCE FROM N.A.
MEDLINE=94042852; PubMed=8226631;
Denome S.A., Stanley D.C., Olson E.S., Young K.D.;
"Metabolism of dibenzothiophene and naphthalene in Pseudomonas
strains: complete DNA sequence of an upper naphthalene catabolic
pathway.";
J. Bacteriol. 175:6890-6901(1993).
-|- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-
CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THBPA).
THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.
-|- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES
CONVERSION OF NAPHTHALENE TO SALICYLATE, AND CATABOLISM OF
DIBENZOTHIOPHENE (DBT) AND PHENANTHRENE. CONVERTS PHENANTHRENE TO
1-HYDROXY-2-NAPHTHOIC ACID AND THE METABOLISM OF DBT IS LIMITED TO
OXIDATION OF THE AROMATIC RING.
-|- MISCELLANEOUS: DOXH AND DOXU ENCODE DIFFERENT ENZYMES THAT MAY
HAVE INTERCHANGEABLE FUNCTIONS.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: M60405; AAA16133.2; ALT_SEQ.
InterPro; IPR004287; HCCA_isomerase.
Pfam; PF03046; HCCA_isomerase; 1.
Isomerase; Pfam10; Aromatic hydrocarbons catabolism.
SEQUENCE 212 AA; 24039 MW; 5EF96A619913DB4F CRC64;
-----
Query Match 7.5%; Score 89; DB 1; Length 212;
Best Local Similarity 24.2%; Pred. No. 0.96;
Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;
YQ 8 VEFYDVLSPYSWLGFEILCRYQNTWNTNLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65
| : : | | | : | : : : : : : : : : : : : : : : : : : : : : : : :
DQ 16 VDFYDFLSPFSYLANHRLSKLAQDYGFSIRYAYDAIDLARVKIAGNVGPSNRDLIVKLDY 75
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YQ 66 MANDKLLRHLHQLPIHPFKDFLSYMLEKSL--SAMRFLTA-VNLEHPMELEKASRELW 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 76 LKVDLQRWAELEYELPVPANYSRRMNTGLYSGAMQATGAYNV-----VF 123
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YQ 123 MRVMSRNEDITEPQSILAA--AEKAGMSAEQAQGLLEKIAETPKVKNLKETTAACRYGA 180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 124 NAVV--GDGIAPDLESPLALVCEKIGWDRSAFE---DFISSDATERVDEQTHAAIERKV 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YQ 181 FGLPTTVAHVQQTHTMLFGSDRMELLAHLGL 211
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 179 FGVP-TMFLGD---EMWGNDRFLFLENVAVG 205
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 8
THD1 ARATH
ID THD1 ARATH STANDARD; PRT; 592 AA.
AC Q92SS6; Q9SPF1;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic, chloroplast precursor
DE (EC 4.2.1.16) (Threonine deaminase) (TD).
GN OM1 OR AT3G10050 OR T22K18.12.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

or send an email to license@isb-sib.ch.

EMBL: AF096281; AAC97936.1; -
 EMBL: AF221984; AAF32370.1; -
 EMBL: AF177212; AAD54324.1; -
 EMBL: AC010927; AAF04418.1; -
 HSP: P04968; ITDJ.
 InterPro: IPR000634; dehydratase_ser_thr.
 InterPro: IPR001926; PALP.
 InterPro: IPR001721; Thr_dehydrat_C.
 Pfam: PF00291; PALP; 1.
 Pfam: PF00585; Thr_dehydrat_C; 2.
 PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
 Isoleucine biosynthesis; Lyase; Chloroplast; Pyridoxal phosphate;
 Allosteric enzyme; Transist peptide.
 TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 CHAIN ? 592 THREONINE DEHYDRATASE BIOSYNTHETIC.
 BINDING 141 141 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 VARIANT 499 499 R -> C (IN STRAIN GML1B).
 VARIANT 544 544 R -> H (IN STRAIN GML1B).
 SEQUENCE 592 AA; 64634 MW; 16658747052FAE7C CRC64;

Query Match 7.5%; Score 89; DB 1; Length 592;
 Best Local Similarity 22.8%; Pred. No. 3.3;
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRLHQLPIHPKDFLSVM-----LEKGSLS-AMRFLT-- 104
 DB 52 PPKLPPLR-----LKVSPNSLQ-----YPAGYLGAVPERTNEANGSTAEAEYLN 99
 QY 105 -----AVNLEHP-EMLEKASRELWVRVMSRNEDITE-----PQSILAA 141
 DB 100 LSTKVYDIATESPLQAKLKLKRLGVRLKREDLPQVFSFKLRGAYNMVMKLPADQAK 159
 QY 142 AEKAGMSAEQAQ-----LLEKATPKVKNKLTETEAACRYGAGLPTIVA 188
 DB 160 GVTCSAGNHAQGVALSASKLCTGTAIVVPTVTPKQAVENL----- 203
 QY 189 HVDGQTHLFGSDRMELLAHL---LGEKWMGP 220
 DB 204 ---GATVWLFQGSYDQAQAKIRAEERGLTFIPP 235

RESULT 9
 SG2_BOVIN
 ID SG2_BOVIN STANDARD; PRT; 613 AA.
 AC P20616;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Secretogranin II precursor (SGII) (Chromogranin C).
 GN SG2 OR CHGC.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9613;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90264409; PubMed=2345170;
 RA Fischer-Colbrie R., Gutierrez J., Hsu C.M., Tacangelo A., Elden L.E.;
 RA "Sequence analysis, tissue distribution and regulation by cell
 RT depolarization, and second messengers of bovine secretogranin II
 RT (chromogranin C) mRNA";
 RL J. Biol. Chem. 265:9208-9213(1990).
 CC -!- FUNCTION: SECRETOGRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE
 CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.
 CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
 CC GRANULES.
 CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
 CC -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN
 CC FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

EMBL: J05468; AAA30760.1; -
 PIR: A35296; A35296.
 InterPro: IPR001990; Granin.
 Pfam: PF01271; Granin; 1.
 PROSITE: PS00422; GRANINS_1; 1.
 KW Sulfation; Cleavage on pair of basic residues; Calcium-binding;
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 30 POTENTIAL.
 FT CHAIN 31 613 SECRETOGRANIN II.
 FT PEPTIDE 181 213 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).
 FT MOD_RES 150 150 SULFATION (BY SIMILARITY).
 SQ SEQUENCE 613 AA; 70356 MW; 5DC079F559D83516 CRC64;

Query Match 7.5%; Score 88.5; DB 1; Length 613;
 Best Local Similarity 23.6%; Pred. No. 3.8;
 Matches 39; Conservative 26; Mismatches 41; Indels 59; Gaps 9;

QY 6 RTVELFVDVLSPYSLWGLFEILLCRYQNIWNI-----NLQLRSLITGIMKDSGNKPPGLLP 60
 DB 333 RAIRLFEPKLPD-----OSIQLIEISRLQIPPELDLMLK-TGEXP--VEP 377
 QY 61 RKGLYMANDLKLRLHQLPIHPKDFLSVMLEKGSLSAMRFTAVNLEHPMLEKASRE 120
 DB 378 EQ-----ELEIPEV-PED-----ISEVDLHPDLFQN----- 403
 QY 121 LWMVRVSRNEDITEPQSILAAAEKAGMSAEQAQGLL--EKIATPK 163
 DB 404 ---KMLSKNGYPRKAPGHAAVEALPEGLSVEDILNLGMSAANPK 445

RESULT 10
 MIAA_TREPA
 ID MIAA_TREPA STANDARD; PRT; 316 AA.
 AC O83644;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE tRNA delta(2)-isopentenylpyrophosphate transferase (EC 2.5.1.8) (IPP
 DE transferase) (Isopentenyl-diphosphate:tRNA isopentenyltransferase)
 DE (IPTase) (IPPT).
 GN MIAA OR TP0637.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OC NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete";
 RL Science 281:375-388(1998).
 CC -!- FUNCTION: CATALYZES THE FIRST STEP IN THE BIOSYNTHESIS OF
 CC 2-METHYLTHIO-N6-(DELTA(2)-ISOPENTENYL)-ADENOSINE (MS[2][6]A)]
 CC ADJACENT TO THE ANTICODON OF SEVERAL tRNA SPECIES (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: isopentenyl diphosphate + tRNA -> diphosphate +

CC crna containing 6-isopentenyladenosine.
 CC -1- SIMILARITY: BELONGS TO THE IPP TRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001238; AAC65611.1; --
 DR TIGR; TP0637; --
 DR InterPro; IPR002627; IPPT.
 DR Pfam; PF01715; IPPT; 1.
 DR ProDom; PD004674; IPPT; 1.
 DR Transfaser; Nucleotidyltransferase; tRNA processing; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 27 34 ATP (POTENTIAL).
 SQ SEQUENCE 316 AA; 36174 MW; E4BE144EE6ED5208 CRC64;
 Query Match 7.1%; Score 84.5; DB 1; Length 316;
 Best Local Similarity 23.3%; Pred. No. 3.8; Mismatches 45; Gaps 8;
 Matches 41; Conservative 30;
 QY 13 DVLSVSWLGFELCRYQNIWNINLQRLPSLITGIMKDSGNKPPGLPRKGLYMANDLKL 72
 DB 80 DCDPVE--EYNVFRFOAVYGI---VPSIL-----RAHKVPVIIVGTGLYLD---AV 124
 QY 73 LRHLQIPIHPKDFLSVLMKSGLSAM-----RELTAVALN- 108
 DB 125 LRQYALVPE-RNQALRASLRGASLSHMRVYFSLKDSHAVHNKTDLEDPARLMRAIEIA 193
 QY 109 ----EHEPMLEKASRELWMYRSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIA 160
 DB 184 VFHATHPELLOQA-RETPRPMRAKVGTYQYPSMLRARIARLEQRIGRGLIEVA 238
 RESULT 11
 S3BL_YEAST
 ID S3BL_YEAST STANDARD; PRT: 971 AA.
 AC P49955;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE U2 snRNP component HSH155.
 GN HSH155 OR YMR288W OR YMR021.14.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: contacts pre-mRNA on both sides of the branch site early
 CC in spliceosome assembly (By similarity).
 CC -1- SUBUNIT: Subunit of the U2 snRNP (By similarity).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SF3B1 FAMILY.
 CC -1- SIMILARITY: CONTAINS 9 HEAT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z49704; CAA89786.1; --
 DR

DR SGD; S0004901; HSH155.
 DR InterPro; IPR000357; HEAT_REPEAT; FALSE_NEG.
 DR PROSITE; PS50077; HEAT_REPEAT; mRNA_splicing; Nuclear protein; Repeat.
 KW Spliceosome; mRNA processing; HEAT 1.
 FT REPEAT 199 237
 FT REPEAT 273 310 HEAT 2.
 FT REPEAT 350 387 HEAT 3.
 FT REPEAT 513 550 HEAT 4.
 FT REPEAT 596 633 HEAT 5.
 FT REPEAT 680 717 HEAT 6.
 FT REPEAT 722 759 HEAT 7.
 FT REPEAT 792 829 HEAT 8.
 FT REPEAT 832 870 HEAT 9.
 SQ SEQUENCE 971 AA; 110027 MW; 27D26EA252A788E2 CRC64;
 Query Match 7.1%; Score 84.5; DB 1; Length 971;
 Best Local Similarity 21.7%; Pred. No. 15;
 Matches 36; Conservative 31; Mismatches 64; Indels 35; Gaps 7;
 QY 5 PRTVEFYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGLPR-KG 63
 DB 387 PYGTEVFNVLPE-----LWKGIRSHRGKVLSSFLKAVGSMPLMDPEYAG 432
 QY 64 LYMANDLKLRRHLLQIPIHPKDFLSVLMKSGLSAMRFLTAVNLEHPEML-EKASRELW 122
 DB 433 YTTTEAMRIIRREFSDPDDEMKKTIILLVLOK--CSAVESIT-----PKFLREETAPEFF 484
 QY 123 MRVSRNEDITEP-----OSILAAAEKAGMSAEQAQGLLEKIATP 162
 DB 485 OKFWRRVVALDRLPKVVTYTTVTLAKKLGCSY-----TIDKLLTP 525
 RESULT 12
 CATA_LISSE
 ID CATA_LISSE STANDARD; PRT: 488 AA.
 AC P24168;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Catalase (EC 1.11.1.6).
 GN KAT.
 OS Listeria seeligeri.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Listeria.
 OX NCBI_TaxID=1640;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91317736; PubMed=1860824;
 RA Haas A., Brehm K., Krefl J., Goebel W.;
 RT "Cloning, characterization, and expression in Escherichia coli of a
 RT gene encoding Listeria seeligeri catalase, a bacterial enzyme highly
 RT homologous to mammalian catalases";
 RL J. Bacteriol. 173:5159-5167(1991).
 CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
 CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
 CC -1- COFACTOR: HEME GROUP.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M75944; AAB53655.1; --
 DR PIR; A40367; A40367.
 DR HSP; P21179; 1CF9.
 DR InterPro; IPR002226; Catalase.
 DR

DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR ProDom: PD000510; Catalase; 1.
 DR PROSITE: PS00437; CATALASE_1; 1.
 DR PROSITE: PS00438; CATALASE_2; 1.
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
 FT ACT_SITE 55
 FT ACT_SITE 128
 FT BINDING 338
 FT BINDING 338
 SQ SEQUENCE 488 AA; 55869 MW; 50A3251469661EBB CRC64;

Query Match
 Best Local Similarity 7.0%; Score 83.5; DB 1; Length 488;
 Matches 52; Conservative 37; Mismatches 76; Indels 79; Gaps 12;

QY 9 ELFDVLPSP-----YSLGFEILCYQNIWNTNLQRPDLITIMKDSGNK 54
 DB DLYVQVLPKLDNDYDFNPLDQKW--FEDVFPYEHVGTMTLNRNPDNIFAETESVGFN 315
 QY 55 P-----PGLLPKRG-----LYMADLKLRL--HHLQIPHFPPKDFLSVMLEKGSLSAMR 101
 DB PGLVPGMLPSDRVLQGRLEFSYDQTHRVGPNYLQPLPNSPKTPVDNNDGDMPPKQ 375
 QY 102 FLTAVNLE-----HPE-----MLEKA-----SRELMMRWVSRNE 130
 DB QTSSINYPNSVDTEPKENPAIYEPQEIIRGDISGRVLAEPKNNFNGHAKVWKR----- 429
 QY 131 DITEPOSTLAAAEKAGM---SAEQAGLLEKIATPKVKNOLKETTEAACRGVAGPLPTV 187
 DB YSDAERAAALVKNIVDDWEGVREDIKIRLNFYQVEPEFAERVA--GTGINL 480
 QY 188 A-HV 190
 DB 481 AEHV 484

RESULT 13
 NCD_DROME
 ID NCD_DROME STANDARD; PRT; 700 AA.
 AC P20480; Q9VAG8;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Claret segregational protein.
 GN NCD OR CA(ND) OR CG7831.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN - [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R, AND CANTON-S; TISSUE=Ovary;
 RX MEDLINE=90231469; PubMed=1691829;
 RA Endow S.A., Henikoff S., Soler-Niedziela L.;
 RT "Mediation of meiotic and early mitotic chromosome segregation in
 Drosophila by a protein related to kinesin."
 RL Nature 345:81-83(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazewicz R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouch J., Brokstein P., Brotter P.,

Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davanport L.B., Davies P.,
 de Fabros B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houtson K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liu X., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE OF 16-700 FROM N.A.
 RX MEDLINE=90275618; PubMed=2140958;
 RA McDonald H.B., Goldstein L.S.B.;
 RT "Identification and characterization of a gene encoding a
 kinesin-like protein in Drosophila."
 RL Cell 61:991-1000(1990).
 RN [4]
 RP MOTOR DIRECTIONALITY.
 RX MEDLINE=91043032; PubMed=2146510;
 RA Walker R.A., Salmon E.D., Endow S.A.;
 RT "The Drosophila claret segregation protein is a minus-end directed
 motor molecule."
 RL Nature 347:780-782(1990).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=94155838; PubMed=8112290;
 RA Lockhart A., Cross R.A.;
 RT "Origins of reversed directionality in the ncd molecular motor."
 RL EMBO J. 13:751-757(1994).
 RN [6]
 RP MUTANT ALLELE NCD(D).
 RX MEDLINE=91122049; PubMed=1825056;
 RA Komma D.J., Horne A.S., Endow S.A.;
 RT "Separation of meiotic and mitotic effects of claret
 non-disjunctional on chromosome segregation in Drosophila."
 RL EMBO J. 10:419-424(1991).
 RN [7]
 RP CHARACTERIZATION OF MUTANT ALLELE NCD(D).
 RX MEDLINE=96283629; PubMed=8670831;
 RA Moore J.D., Song H., Endow S.A.;
 RT "A point mutation in the microtubule binding region of the ncd motor
 protein reduces motor velocity."
 RL EMBO J. 15:3306-3314(1996).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 335-700.
 RX MEDLINE=96195067; PubMed=8606780;
 RA Sablin E.P., Kull F.J., Cooke R., Vale R.D., Fletcher R.J.;
 RT "Crystal structure of the motor domain of the kinesin-related motor
 ncd."
 RL Nature 380:555-559(1996).
 CC - FUNCTION: NCD IS REQUIRED FOR NORMAL CHROMOSOMAL SEGREGATION IN
 MEIOSIS, IN FEMALES, AND IN EARLY MITOTIC DIVISIONS OF THE EMBRYO.
 CC "THE NCD MOTOR ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S MINUS
 END."

Tue May 21 08:04:02 2002

```

CC CC MEDLINE-8909327; PubMed-32111750;
CC RA Gerdes H.-H., Philipps E., Huttner W.B.:
CC RT "The primary structure of rat secretogranin II deduced from a cDNA
CC sequence."
CC RL Nucleic Acids Res. 16:11811-11811(1988).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC MEDLINE-93309708; PubMed-8321414;
CC RA Kakar S.-S., Wei N., Mulchahey J.J., Leboeuf R.D., Neill J.D.:
CC RT "Regulation of expression of secretogranin II mRNA in female rat
CC pituitary and hypothalamus."
CC RL Neuroendocrinology 57:422-431(1993).
CC CC -1- FUNCTION: SECRETOGHRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE
CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.
CC CC -1- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
CC GRANULES.
CC CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC CC -1- SIMILARITY: BELONGS TO THE CHROMOGHRANIN / SECRETOGHRANIN PROTEIN
CC FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC EMBL; X13618; CAA31950.1; -
CC DR EMBL; M93669; AAM42135.1; -
CC DR PIR; S02180; S02180.
CC DR InterPro; IPR001990; Granin.
CC DR Pfam; PF01271; Granin; 1.
CC DR PROSITE; PS00422; GRANINS_1; 1.
CC KW Sulfation; Cleavage on pair of basic residues; Calcium-binding;
CC Signal.
CC FT SIGNAL. 1 30 SECRETOGHRANIN II.
CC FT CHAIN 31 619 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).
CC FT PEPTIDE 184 216 SULFATION (BY SIMILARITY).
CC FT MOD_RES 153 153
CC SQ SEQUENCE 619 AA; 71031 MW; 27CB75B4F25A38D1 CRC64;

Query Match 6.9%; Score 81.5; DB 1; Length 619;
Best Local Similarity 24.8%; Pred. No. 16;
Matches 41; Conservative 16; Mismatches 57; Indels 51; Gaps 8;

QY 13 DVLSYSLWGLFEILCRYQNIWNINQLRPSLITIGMKDSGNKPPGLPRKGLYMANDKL 72
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 105 DVLSEDEWM-----RIIEALRQAEENPPSALKENKPYALNLEK- 143
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 73 LRHLQIPHPKDFLSVMLEKGSLSAMRFLTAVNLEHPMLEKASRELMMRVMSRVEDI 132
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 144 -----NPFVDPDDYETQOWPERKLEKMRP-----PLMYEENSRE----NPFKRTNEI 187
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 133 TE-----PQSILAAAEKAGMSABQAGLLEKIATPKVKNQLKETTE 173
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 188 VEEQYTPQS-LATLESVFOE-----LGKLTGP--SNQKRERVD 222
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
DPOI_ADE12 STANDARD; PRT; 1061 AA.
ID DPOI_ADE12
AC P05538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL.
OS Human adenovirus type 12.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28282;
RN [1]

```

```

CC CC MEDLINE-8909327; PubMed-32111750;
CC RA Gerdes H.-H., Philipps E., Huttner W.B.:
CC RT "The primary structure of rat secretogranin II deduced from a cDNA
CC sequence."
CC RL Nucleic Acids Res. 16:11811-11811(1988).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC MEDLINE-93309708; PubMed-8321414;
CC RA Kakar S.-S., Wei N., Mulchahey J.J., Leboeuf R.D., Neill J.D.:
CC RT "Regulation of expression of secretogranin II mRNA in female rat
CC pituitary and hypothalamus."
CC RL Neuroendocrinology 57:422-431(1993).
CC CC -1- FUNCTION: SECRETOGHRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE
CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.
CC CC -1- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
CC GRANULES.
CC CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC CC -1- SIMILARITY: BELONGS TO THE CHROMOGHRANIN / SECRETOGHRANIN PROTEIN
CC FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC EMBL; X52814; CAA36998.1; -
CC DR EMBL; M33932; AAA28716.1; -
CC DR EMBL; AE003771; AAF56942.1; -
CC DR EMBL; X57475; CAA40713.1; -
CC DR PIR; A35624; A35624.
CC DR PIR; S09748; S09748.
CC DR HSP; P17119; 3KAR.
CC DR FlyBase; FBgn0002924; ncd.
CC DR InterPro; IPR001752; kinesin.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PR00380; KINESINHEAVY.
CC DR SMART; SM00129; KTSC; 1.
CC DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
CC KW Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
CC Meiosis; Mitosis.
CC FT DOMAIN 196 346 COILED COIL (POTENTIAL).
CC FT DOMAIN 347 700 KINESIN-MOTOR (BY SIMILARITY).
CC FT NP_BIND 434 441 ATP (BY SIMILARITY).
CC FT MOTAGEN 556 556 V->F; IN NCD(D); REDUCES MOTOR VELOCITY.
CC FT CONFLICT 697 697 S -> N (IN REF. 1).
CC SQ SEQUENCE 700 AA; 77473 MW; ADE043CBCE7FD561 CRC64;

Query Match 7.0%; Score 83.5; DB 1; Length 700;
Best Local Similarity 22.4%; Pred. No. 12;
Matches 52; Conservative 37; Mismatches 70; Indels 73; Gaps 12;

QY 1 MGPLPRTVELFDVLSYSLWGL--FEILCRYQNIWNINQLRPSLITIGMKD-----SG 52
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 451 VGVIPRTVDLFDSDRGYNLGNWEYELKATFLEIYN---EVLYDLLSNEQKDMIRMAKN 507
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 53 NKPPCLLPKGLYMAN---DLKLLRHHLQIPHPKDFLSVMLEKGSLSAMRFLTAVNLE 109
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 508 NK-----NDIYVSNITEETVLDPNHLRLHMTAK-----MNRATAS---TAGN-- 547
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 HPMELEKASRELMMRVMSRVEDITEPQSILAAAEKAGMS-----AEQAQGLLEK 158.
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 548 -----ERS-----SRSHAVTKLEIGRHAQKQISVGSINLVDLAGSESPKSTRM 593
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 159 IATPKVKNLKETTEAACRYGAFGLPTTVAHVQGTMLFGSDRMELLALLH 210
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 594 TETKNINRSLSELTN-----VILALLQKDHIPY---RNSKLTHLL 631
||| | : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
SG2_RAT STANDARD; PRT; 619 AA.
ID SG2_RAT
AC P10362;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Secretogranin II precursor (SGII) (Chromogranin C).
GN SG2 OR CHGC OR SCG-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=94076430; PubMed=8254750;
RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;
RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
RL functional analysis";
RN J. Virol. 68:379-389(1994).
RP SEQUENCE FROM N.A.
RX MEDLINE=87106854; PubMed=3803925;
RA Shu L., Hong J.S., Wei Y.-F., Engler J.A.;
RT "Nucleotide sequence of the genes encoded in early region 2b of human
RL adenovirus type 12";
CC Gene 46:187-195(1986).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73487; CAA51882.1; -
DR EMBL; M14785; AAA2478.1; ALT_INIT.
DR PIR; A25770; DJAD12.
DR PIR; S33933; S33933.
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
FT CONFLICT 32 32 R -> S (IN REF. 2).
FT CONFLICT 162 162 V -> L (IN REF. 2).
FT CONFLICT 181 182 LQ -> YN (IN REF. 2).
FT CONFLICT 461 461 S -> T (IN REF. 2).
FT CONFLICT 575 575 L -> F (IN REF. 2).
FT CONFLICT 892 892 S -> T (IN REF. 2).
FT CONFLICT 1030 1030 K -> M (IN REF. 2).
SQ SEQUENCE 1061 AA; 121172 MW; 33FBA89C33065C08 CRC64;

Query Match
Best Local Similarity 6.9%; Score 81.5; DB 1; Length 1061;
Matches 50; Conservative 35; Mismatches 77; Indels 67; Gaps 12;
Qy 1 MGPLPRTVELF--YDVLSPYSWLGFEILCRYQNIWNINLQRPISLITGIMKDSGNKPPGL 58
Db 133 IGSHPRTERLFTYDV-ETVTWNG-----AFGKQLVPEML--VMKLSG----- 172
Qy 59 LPRKGLYMANDKLRLHHLQIPI-----HPPKO---FLSVMLEKGSLSAMRFLTAVNLEH 110
Db 173 -----DNLVYKHALQALALELQWDQWEKOSTTEYCLTPKMKV-GQQFRTYRN--- 218
Qy 111 PEMLEKASRELNRVWSNEDITEPOSILAAAEKAGMSAQOGLLEKIATPKVKNOLKE 170
Db 219 -RLQTSLATDLMMTFLQKNPHLSQ-----WAQENGIVALEDLSEYDLKRAPAIKGEPR- 271
Qy 171 TTEACRYGAFGLPITVAHVGDGTHMFGSDRMELLALLGKWMGPIP 219
Db 272 -----FVELYIVG--HNINGFDEIVLAAQVINNRDVPGP 304

```

Search completed: May 20, 2002, 08:42:28
Job time: 195 sec

Tue May 21 08:04:00 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:33:28 ; Search time 30.57 Seconds
(without alignments)
821.154 Million cell updates/sec

Title: US-09-441-723-1
Perfect score: 1185
Sequence: 1 MGPLPRTVELFDVLSYSW.....AHLIGKWMGPPIPPAVNARL 226

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 747574 seqs, 111073796 residues 747574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match %	Length	ID	Description
1	1185	100.0	226	AA1980	Novel human glutathione S-transferase
2	1185	100.0	226	AA1981	Human endometrium
3	820.5	69.2	300	AA1982	Human OREF 3062
4	700	59.1	131	AA1983	Human secreted pro
5	496	41.9	97	AA1984	Human secreted pro
6	387	32.7	97	AA1985	Gene 45 human secr
7	326.5	27.6	107	AA1986	Human reproductive
8	89.5	7.6	1086	AA1987	Novel human diago
9	89	7.5	545	AA1988	Mutant threonine d
10	89	7.5	545	AA1989	Feedback insensiti
11	89	7.5	590	AA1990	Mutant threonine d

12	89	7.5	592	20	AA1991	Mutant threonine d
13	89	7.5	592	20	AA1992	Mutant threonine d
14	89	7.5	592	20	AA1993	Wild type threonin
15	89	7.5	592	20	AA1994	Arabidopsis wild-t
16	89	7.5	592	20	AA1995	Feedback insensiti
17	89	7.5	592	20	AA1996	Feedback insensiti
18	89	7.5	592	20	AA1997	Mutant threonine d
19	89	7.5	592	20	AA1998	Feedback insensiti
20	89	7.5	592	20	AA1999	Mutant threonine d
21	89	7.5	592	20	AA2000	Feedback insensiti
22	89	7.5	592	20	AA2001	Mutant threonine d
23	89	7.5	592	20	AA2002	Feedback insensiti
24	89	7.5	592	20	AA2003	Mutant threonine d
25	89	7.5	592	20	AA2004	Feedback insensiti
26	89	7.5	592	20	AA2005	Mutant threonine d
27	89	7.5	592	20	AA2006	Feedback insensiti
28	89	7.5	592	20	AA2007	Mutant threonine d
29	89	7.5	592	20	AA2008	Feedback insensiti
30	89	7.5	592	20	AA2009	Mutant threonine d
31	89	7.5	592	20	AA2010	Feedback insensiti
32	89	7.5	592	20	AA2011	Mutant threonine d
33	89	7.5	592	20	AA2012	Feedback insensiti
34	89	7.5	592	20	AA2013	Mutant threonine d
35	89	7.5	592	20	AA2014	Feedback insensiti
36	89	7.5	592	20	AA2015	Mutant threonine d
37	89	7.5	592	20	AA2016	Feedback insensiti
38	89	7.5	592	20	AA2017	Mutant threonine d
39	89	7.5	592	20	AA2018	Feedback insensiti
40	89	7.5	592	20	AA2019	Mutant threonine d
41	89	7.5	592	20	AA2020	Feedback insensiti
42	89	7.5	592	20	AA2021	Mutant threonine d
43	89	7.5	592	20	AA2022	Feedback insensiti
44	89	7.5	592	20	AA2023	Mutant threonine d
45	89	7.5	592	20	AA2024	Feedback insensiti

ALIGNMENTS

RESULT 1
AA1991
ID AA1991 standard; Protein; 236 AA.
XX
AC AA1991
XX
DT 05-JUN-2000 (first entry)
XX
DE Novel human glutathione S-transferase, GSTS.
XX
KW Glutathione S-transferase; human; GSTS; cancer; immune disorder;
KW gene therapy; diagnosis; treatment; drug screening.
XX
OS Homo sapiens.
XX
PN US6030809-A.
XX
PD 29-FEB-2000.
XX
PF 25-NOV-1997; 97US-0978174.
XX
PR 25-NOV-1997; 97US-0978174.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Shah P, Lal P, Corley NC;
XX
DR WPI; 2000-205204/18.
XX
DR N-PSDB; AA202599.
XX
PT Isolated nucleic acid encoding glutathione S-transferase useful in the
PT production of agents for preventing, diagnosing and treating diseases
PT associated with cell proliferation -

PS This sequence represents a novel human glutathione S-transferase, GSTS.
XX Nucleotide sequences encoding GSTS were initially isolated from a
CC urologic CDNA library, and subsequently extended using cDNA libraries
CC derived from other tissues, such as brain or bladder. The present
CC sequence is encoded by a consensus cDNA, GSTS, and nucleotides which
CC encode it may be used in the prevention, GSTS, and nucleotides which
CC diseases associated with inappropriate GSTS expression, such as cancers
CC and immune disorders. Nucleotides which encode GSTS may be used in gene
CC therapy to treat disorders associated with reduced expression or activity
CC of GSTS, and in antisense therapy for disorders associated with increased
CC GSTS expression or activity. They may also be used for the recombinant
CC production of GSTS, and as a source of probes and primers to detect and
CC quantitate the presence of similar nucleic acid sequences, particularly
CC for the diagnosis of GSTS-associated disorders. GSTS proteins may be used
CC as antigens in the production of antibodies against GSTS and in assays to
CC identify modulators (agonists and antagonists) of GSTS expression and
CC activity. The anti-GSTS antibodies and GSTS antagonists may also be used
CC to downregulate GSTS expression and activity. Antagonists of GSTS
CC expression and function may be used to treat immune disorders (e.g., AIDS,
CC anaemia, asthma, Crohn's disease, irritable bowel syndrome, multiple
CC sclerosis, osteoarthritis, microbial infections) and cancers (e.g.,
CC leukaemia, lymphoma, melanoma, and cancers of the breast, prostate,
CC liver, lung and brain). The anti-GSTS antibodies may also be used as
XX diagnostic agents.
SQ Sequence 226 AA;

Query Match 100.0%; Score 1185; DB 21; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.3e-121;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60
DB 1 mgplprtvelfdvlsyswlgfeilcryqniwnlnqlrpslittgimkdsngkppglip 60
QY 61 RKGLYMANDKLLRHHLQIPIHFPKDFLSVLEKGSLSAMRFLTAVNLEHPMLEKASRE 120
DB 61 rkglymandkllrhhllqpihfpkdfslvmlkgsisamrfltaavnlehpemlekasre 120
QY 121 LWMRVWSRNEDITEPOSILAAAEKAGMSAEQAGGILLEKIATPKVKNQKETTAAACRYGA 180
DB 121 lwmrwvwsrnediteposilaaekagmsaeqaggllekiatpkvknqketteaacryga 180
QY 181 FGLPITVAHVQDQTHMLFGSDRMELLAHLLGKWMGPIPPAVNARL 226
DB 181 fglpityvahvdqthmlfgsdrmlahllgkwmgpippavnarl 226

RESULT 2
AA59988
ID AAY59988 standard; Protein; 256 AA.
XX AC AAY59988;
XX AC AAY59988;
XX 31-JAN-2000 (first entry)
XX Human endometrium tumour EST encoded protein 48.
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
KW treatment; uterine; gene therapy; expressed sequence tag.
XX Homo sapiens.
XX DE19817948-A1.
XX 21-OCT-1999.
XX 17-APR-1998; 98DB-1017948.
XX 17-APR-1998; 98DB-1017948.
PR 17-APR-1998; 98DB-1017948.

XX
PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
XX WPI; 1999-591957/51.
XX N-PSDB; AA241996.
XX New nucleic acid sequences expressed in uterine cancer tissues, and
XX derived polypeptides for treatment of uterine and endometrial cancer
XX and identification of therapeutic agents -
XX Claim 23; Page 294; 444pp; German.
XX This invention describes novel human nucleic acid (cDNA) sequences (A),
XX that are highly expressed in uterine tumour tissue and which have
XX anticancer and cytostatic activity. (A) are used (i) for recombinant
XX expression of polypeptides (B) and (ii) to isolate complete genes. (B)
XX are used (i) to identify agents suitable for treatment of uterine or
XX endometrial cancer; (ii) directly for treating these forms of cancer
XX (including expression from gene therapy vectors) and (iii) for
XX generation of specific antibodies. (A) are identified by assembling ESTs
XX (expressed sequence tags) from a particular tissue type before comparison
XX of expression patterns. This allows a significantly longer fragment of
XX the gene to be revealed, so should reduce the number of failures
XX associated with the fact that ESTs from different libraries may represent
XX different parts of the same unknown gene, distorting the estimated
XX frequency of occurrence in a particular tissue. AAY59941-Y60328 represent
XX protein fragments encoded by the human endometrium tumour cDNA library
XX derived EST fragments represented in AA241981-242121.
SQ Sequence 256 AA;

Query Match 100.0%; Score 1185; DB 20; Length 256;
Best Local Similarity 100.0%; Pred. No. 1.6e-121;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60
DB 31 mgplprtvelfdvlsyswlgfeilcryqniwnlnqlrpslittgimkdsngkppglip 90
QY 61 RKGLYMANDKLLRHHLQIPIHFPKDFLSVLEKGSLSAMRFLTAVNLEHPMLEKASRE 120
DB 91 rkglymandkllrhhllqpihfpkdfslvmlkgsisamrfltaavnlehpemlekasre 150
QY 121 LWMRVWSRNEDITEPOSILAAAEKAGMSAEQAGGILLEKIATPKVKNQKETTAAACRYGA 180
DB 151 lwmrwvwsrnediteposilaaekagmsaeqaggllekiatpkvknqketteaacryga 210
QY 181 FGLPITVAHVQDQTHMLFGSDRMELLAHLLGKWMGPIPPAVNARL 226
DB 211 fglpityvahvdqthmlfgsdrmlahllgkwmgpippavnarl 256

RESULT 3
AAB43298
ID AAB43298 standard; Protein; 300 AA.
XX AC AAB43298;
XX AC AAB43298;
XX 08-FEB-2001 (first entry)
XX Human ORF3062 polypeptide sequence SEQ ID NO:6124.
XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
KW vulnary; antiprosclatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antineoplastic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

Tue May 21 08:04:00 2002

cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI: 2000-602362/57.

N-PSDB; AAC77507.

Novel nucleic acids and peptides derived from open reading frame X,

useful for treating e.g. cancers, proliferative disorders,

neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 5304-5305; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, lupus erythematosus, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 300 AA;

Query Match 59.2%; Score 820.5; DB 21; Length 300;

Best Local Similarity 72.08; Pred. No. 2e-81;

Matches 167; Conservative 3; Mismatches 1; Indels 61; Gaps 2;

QY 41 PSLITGIMKDSGNKPPGILLPRKGLYMANDLKLRHLLQIPHPKDFLSVMLEKGSLSAM 100

Db 74 pnlll-----agnkppgllprkgllymandklrrhllqipihfpkdfslvmlekgslsam 128

QY 101 RFLFVNLVHEPMELEKASRELWVRWSR-----NEDITEPQSILAAAEKAGMSAEOAGLLEKIATPKV 164

Db 129 rfltavnlehpemlekasrelwvrwsrvsvglwessgrtdldflfprhvrvmllppp 188

QY 129 -----NEDITEPQSILAAAEKAGMSAEOAGLLEKIATPKV 164

Db 189 ggstvlvptlpsrlpavfsssqnditepqsilaaekagsaqaqglllekiatpkv 248

QY 165 KNQLKETTEACRYGAFGLPITVAHVVDGQTHMLFGSDRMELLALLGKWMG 216

Db 249 knqlketteacrygafgipitvahvdgqthmlfgsdrmelallhllgkwmg 300

RESULT 4

AAG01523

ID AAG01523 standard; Protein; 131 AA.

XX AAG01523;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 5604.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX N-PSDB; AAC01529.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 13; SEQ ID 5604; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 131 AA;

Query Match 59.1%; Score 700; DB 21; Length 131;

Best Local Similarity 100.0%; Pred. No. 9.9e-69;

Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLRTEVELFVDVLSYSPSWLGFELICRYQNIWININQLRPSLITGIMKDSGNKPPGLLP 60

Db 1 mgplrtvelfydvlsypswlgyfelicryqniwininqlrpslittgimkdsgnkppgllp 60

QY 61 RGLYMANDLKLRHLLQIPHPKDFLSVMLEKGSLSAMRFLFVNLVHEPMELEKASRE 120

Db 61 rkglymandklrrhllqipihfpkdfslvmlekgslsamrflfavnlehpemlekasre 120

QY 121 LWMRVWSRNEED 131

XXXXXXXXXX

Db 121 lmrwvwrned 131

RESULT 5
ID AAB34972
XX AAB34972 standard; Protein; 97 AA.
AC AAB34972;
XX
XX
XX 26-JAN-2001 (first entry)
XX Human secreted protein sequence encoded by gene 45 SEQ ID NO:176.
XX Human; secreted protein; neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;
XX antinefactive; gynaecological; antibacterial; neural disorder; cancer;
XX immune disease; reproductive disorder; proliferative disease;
XX gastrointestinal disease; wound healing; infectious disease;
XX food additive.
XX Homo sapiens.
XX OS
XX PN W0200056766-A1.
XX PD 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06824.
XX
XX 19-MAR-1999; 99US-0125359.
XX 03-DEC-1999; 99US-0168664.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-594574/56.
XX N-PSDB; AAC60069.
XX
XX Human secreted proteins and gene sequences encoding them, useful for
XX detection, prevention, and treatment of various disorders such as
XX cancer and immune system disorders -
XX
XX Disclosure; Page 83; 442pp; English.
XX
XX The polynucleotide sequences given in AAC60025-C60071 encode the human
XX secreted proteins represented in AAB34854-B34900. Sequences
XX AAB34901-B34976 are fragments of proteins encoded by the genes, and also
XX proteins with which they share sequence homology. The proteins have
XX activities based on the tissues in which their encoding genes are
XX expressed. Examples of the proteins activities include: neuroprotective;
XX - cytostatic; cardioactive; immunomodulatory; general muscular activity;
XX antinefactive; gynaecological; and antibacterial. The human secreted
XX proteins, polynucleotides, antagonists and antagonists of the invention
XX may be useful in treating, preventing and/or diagnosing various
XX diseases, disorders and conditions such as neural, immune, muscular,
XX proliferative disorders and cancer. They may also be used in the
XX treatment of wounds, and infectious diseases. The polypeptides may be
XX used as a food additive or preservative to increase storage capabilities.
XX Sequences AAC60016-C60024 and AAB34853 are used in the course of the
XX invention during the identification and characterisation of the protein
XX and nucleotide sequences.
XX
XX Sequence 97 AA;
XX
Query Match 41.9%; Score 496; DB 21; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.6e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 EDITEPQSILAAAEKAGMSAEQAQGLLEKATPKVKNQKLTTEAACRYGAFGLPITVAH 189
|||||

Db 1 editpeqsilaaekagmsaeqaqgllekiaatkpkvknqkltteaacrygafglpitvah 60

QY 190 VDGQTHMFGSDRMELLAHLGKWKMGPIPPAVNARL 226
|||||

Db 61 vdgqthmfgsdrrmellhllgkwmgpiipavnarl 97
|||||

RESULT 6
AAB34971
ID AAB34971 standard; Protein; 97 AA.
XX
XX AAB34971;
XX
XX 26-JAN-2001 (first entry)
XX
XX Gene 45 human secreted protein homologous amino acid sequence #175.
XX Human; secreted protein; neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;
XX antinefactive; gynaecological; antibacterial; neural disorder; cancer;
XX immune disease; reproductive disorder; proliferative disease;
XX gastrointestinal disease; wound healing; infectious disease;
XX food additive.
XX Rattus sp.
XX OS
XX PN W0200056766-A1.
XX PD 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06824.
XX
XX 19-MAR-1999; 99US-0125359.
XX 03-DEC-1999; 99US-0168664.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-594574/56.
XX
XX Human secreted proteins and gene sequences encoding them, useful for
XX detection, prevention, and treatment of various disorders such as
XX cancer and immune system disorders -
XX
XX Disclosure; Page 83; 442pp; English.
XX
XX The polynucleotide sequences given in AAC60025-C60071 encode the human
XX secreted proteins represented in AAB34854-B34900. Sequences
XX AAB34901-B34976 are fragments of proteins encoded by the genes, and also
XX proteins with which they share sequence homology. The proteins have
XX activities based on the tissues in which their encoding genes are
XX expressed. Examples of the proteins activities include: neuroprotective;
XX - cytostatic; cardioactive; immunomodulatory; general muscular activity;
XX antinefactive; gynaecological; and antibacterial. The human secreted
XX proteins, polynucleotides, antagonists and antagonists of the invention
XX may be useful in treating, preventing and/or diagnosing various
XX diseases, disorders and conditions such as neural, immune, muscular,
XX proliferative disorders and cancer. They may also be used in the
XX treatment of wounds, and infectious diseases. The polypeptides may be
XX used as a food additive or preservative to increase storage capabilities.
XX Sequences AAC60016-C60024 and AAB34853 are used in the course of the
XX invention during the identification and characterisation of the protein
XX and nucleotide sequences.
XX
XX Sequence 97 AA;
XX

Query Match 32.7%; Score 387; DB 21; Length 97;
Best Local Similarity 76.3%; Pred. No. 1.4e-34;
Matches 74; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

us-09-441-723-1.1.rag

Tue May 21 08:04:00 2002

QY 130 EDITEPQSTILAAAEKAGMSAQGLLEKATPKVKNKOLKETTAAACRYGAFGLPITVAH 189
 Db 1 editesqnlisaaekagmagaqahllnkistelvkskirettgaackyygaiglpittvah 60
 QY 190 VQGTMLFSGSDRMELLAHLIGKWMGPIPPAVNARL 226
 Db 61 vdgktymlfsgsdrmellaylllgkwmgpvptlnarl 97

RESULT 7
 AAM96474
 ID AAM96474 standard; Protein; 107 AA.
 XX *
 AC AAM96474;
 DT 21-NOV-2001 (first entry)
 DE - Human reproductive system related antigen SEQ ID NO: 5132.
 XX - Human reproductive system related antigen; reproductive system disorder;
 KW Human; reproductive system related antigen; cancer; gene therapy.
 KW cancer; gene therapy.
 XX Homo sapiens.
 OS WO200155320-A2.
 PN 02-AUG-2001.
 PD 17-JAN-2001; 2001NO-US01339.
 PF 31-JAN-2000; 2000US-0179055.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 28-JUN-2000; 2000US-0209467.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216680.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 14-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234274.
 PR 21-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 25-SEP-2000; 2000US-0234984.
 PR 26-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.

[illegible]

QY 142 AERAGMSAEQAQ-----LLEKIATPKVKNOLKETTEACRYGAFGLPTVA 188
 Db 113 gvicssagnhaqgvalsasklgctavimpvttpeikwqavlenl----- 156
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220
 Db 157 ---gatvvlfqdsydaqahakiraeeegltfipp 188

RESULT 11
 AAY32950
 ID AAY32950 standard; Protein; 590 AA.
 AC AAY32950;
 XX 09-NOV-1999 (first entry)
 DE Mutant threonine dehydratase/deaminase protein sequence.

Threonine dehydratase/deaminase; TD; feedback insensitive mutant;
 molecular marker; isoleucine toxic structural analog resistance;
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
 KW polyhydroxybutyrate; antibiotic resistance marker; mutein.
 XX Arabidopsis thaliana.
 OS Synthetic.

PN WO9941395-A1.
 XX 19-AUG-1999.
 PD 08-JAN-1999; 99WO-US00560.
 PF 10-JUL-1998; 98WO-US14362.
 XX 17-FEB-1998; 98US-0074875.
 PR (DOWC) DOW AGROSCIENCES LLC.
 PA (PURD) PURDUE RES FOUND.

PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;
 DR WPI; 1999-527375/44.
 DR N-PSDB; AA211208.

PS New nucleic acid encoding threonine dehydratase deaminase resistant
 PT to feedback inhibition, useful as selection marker for cell
 PT transformation and to impart herbicide resistance
 XX Example 3; Page 119-123; 194pp; English.

CC This sequence represents a mutant Arabidopsis thaliana threonine
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular
 CC marker (impacting resistance to toxic structural analogues of isoleucine)
 CC for selecting transformed cells and to produce transformants with
 CC increased levels of isoleucine (and thus better nutritional value) or of
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also
 CC TD-expressing plants permit use of the isoleucine structural analogues as
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance
 CC markers (which are potentially harmful to the environment). Since no
 CC human analog of TD exists (humans can not synthesize isoleucine), it
 CC should be safe to use.

XX Sequence 590 AA;

Query Match 7.58; Score 89; DB 20; Length 590;
 Best Local Similarity 22.8%; Pred. No. 1.1;
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;
 QY 55 PPGCL-LPRKGLYMANDLKLRHHLQIPHFPPKDFLSVM-----LEKGSLS-AMRFLT-- 104

Db 50 ppklplpr-----lkvspnslq---ypagylqavpertenaeagsiaameyltnl 97
 QY 105 -----AVNLEHP-EMLEKASRELWMRVWSRNETITE-----PQSILAA 141
 Db 98 lstkvydiaiesplqiaakkskrlgvrnylkrredlqpvsfklrgaynmvmkllpadqlak 157
 QY 142 AERAGMSAEQAQ-----LLEKIATPKVKNOLKETTEACRYGAFGLPTVA 188
 Db 158 gvicssagnhaqgvalsasklgctavimpvttpeikwqavlenl----- 201
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220
 Db 202 ---gatvvlfqdsydaqahakiraeeegltfipp 233

RESULT 12
 AAY32939
 ID AAY32939 standard; Protein; 592 AA.
 AC AAY32939;
 XX 09-NOV-1999 (first entry)
 DE Mutant threonine dehydratase/deaminase protein sequence.

Threonine dehydratase/deaminase; TD; feedback insensitive mutant;
 molecular marker; isoleucine toxic structural analog resistance;
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
 KW polyhydroxybutyrate; antibiotic resistance marker; mutein.
 XX Arabidopsis thaliana.
 OS Synthetic.

PN WO9941395-A1.
 XX 19-AUG-1999.
 PD 08-JAN-1999; 99WO-US00560.
 PF 10-JUL-1998; 98WO-US14362.
 XX 17-FEB-1998; 98US-0074875.
 PR (DOWC) DOW AGROSCIENCES LLC.
 PA (PURD) PURDUE RES FOUND.

PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;
 DR WPI; 1999-527375/44.
 DR N-PSDB; AA211197.

PS New nucleic acid encoding threonine dehydratase deaminase resistant
 PT to feedback inhibition, useful as selection marker for cell
 PT transformation and to impart herbicide resistance
 XX Claim 13; Fig 8; 194pp; English.

CC This sequence represents a mutant Arabidopsis thaliana threonine
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular
 CC marker (impacting resistance to toxic structural analogues of isoleucine)
 CC for selecting transformed cells and to produce transformants with
 CC increased levels of isoleucine (and thus better nutritional value) or of
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also
 CC TD-expressing plants permit use of the isoleucine structural analogues as
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance
 CC markers (which are potentially harmful to the environment). Since no
 CC human analog of TD exists (humans can not synthesize isoleucine), it
 CC should be safe to use.

XX Sequence 592 AA;

Tue May 21 08:04:00 2002

CC markers (which are potentially harmful to the environment). Since no
CC human analog of TD exists (humans can not synthesize isoleucine), it
CC should be safe to use.
XX Sequence 592 AA;

Query Match 7.5%; Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRRHQLQIPHFDPFLSV-----LEKGSLS-AMRELT-- 104
Db 52 pklplpr-----lkvspnslq-----ypagyigavpntneangsaameyltni 99
QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----PQSILAA 141
Db 100 lskvydiaiesplqlaklskrigvmylkredlqpvfsklrgaynmvmkkipadqlak 159
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188
Db 160 gvicssagnhagvalsasklgctavimpvttpeikwqavnl----- 203
QY 189 HVDGQTHMFGSDRMELLAHL---LGEKWMGPPIP 220
Db 204 ---gatvvifgdsydaqahakiraeegltfipp 235

RESULT 13
AA32948 standard; Protein; 592 AA.

XX AC AAY32948;
XX DT 09-NOV-1999 (first entry)
XX DE Mutant threonine dehydratase/deaminase protein sequence.
XX KW Threonine dehydratase/deaminase; TD: feedback insensitive mutant;
XX KW molecular marker; isoleucine toxic structural analog resistance;
XX KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
XX KW polyhydroxybutyrate; antibiotic resistance marker; mutin.

XX OS Arabidopsis thaliana.
XX OS Synthetic.
XX FN WO9941395-A1.
XX PD 19-AUG-1999.
XX PF 08-JAN-1999; 99WO-US00560.
XX PR 10-JUL-1998; 98WO-US14362.
XX PR 17-FEB-1998; 98US-0074875.
XX PA (DOWC) DOW AGROSCIENCES LLC.
XX PA (PURD) PURDUE RES FOUND.

XX PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;
XX WPI; 1999-527375/44.
XX DR N-PSDB; AAZ11206.

XX New nucleic acid encoding threonine dehydratase/deaminase resistant
XX to feedback inhibition, useful as selection marker for cell
XX transformation and to impart herbicide resistance
XX Example 3; Page 126-129; 194pp; English.

XX This sequence represents a mutant Arabidopsis thaliana threonine
XX dehydratase/deaminase (TD) protein of the invention. The protein is a
XX feedback insensitive mutant. The TD DNA sequence is used as molecular
XX marker (imparting resistance to toxic structural analogues of isoleucine)
XX for selecting transformed cells and to produce transformants with
XX increased levels of isoleucine (and thus better nutritional value) or of
XX intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for
XX synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also
XX TD-expressing plants permit use of the isoleucine structural analogues as
XX herbicides. The DNA sequences are alternatives for antibiotic resistance

Query Match 7.5%; Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;
QY 55 PPGL-LPRKGLYMANDLKLRRHQLQIPHFDPFLSV-----LEKGSLS-AMRELT-- 104
Db 52 pklplpr-----lkvspnslq-----ypagyigavpntneangsaameyltni 99
QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----PQSILAA 141
Db 100 lskvydiaiesplqlaklskrigvmylkredlqpvfsklrgaynmvmkkipadqlak 159
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188
Db 160 gvicssagnhagvalsasklgctavimpvttpeikwqavnl----- 203
QY 189 HVDGQTHMFGSDRMELLAHL---LGEKWMGPPIP 220
Db 204 ---gatvvifgdsydaqahakiraeegltfipp 235

RESULT 14
AA32951 standard; Protein; 592 AA.

XX AC AAY32951;
XX DT 09-NOV-1999 (first entry)
XX DE Wild type threonine dehydratase/deaminase protein sequence.
XX KW Threonine dehydratase/deaminase; TD: feedback insensitive mutant;
XX KW molecular marker; isoleucine toxic structural analog resistance;
XX KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
XX KW polyhydroxybutyrate; antibiotic resistance marker.

XX OS Arabidopsis thaliana.

XX FN WO9941395-A1.
XX PD 19-AUG-1999.
XX PF 08-JAN-1999; 99WO-US00560.
XX PR 10-JUL-1998; 98WO-US14362.
XX PR 17-FEB-1998; 98US-0074875.
XX PA (DOWC) DOW AGROSCIENCES LLC.
XX PA (PURD) PURDUE RES FOUND.

XX PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;
XX WPI; 1999-527375/44.
XX DR N-PSDB; AAZ11209.

XX New nucleic acid encoding threonine dehydratase/deaminase resistant
XX to feedback inhibition, useful as selection marker for cell
XX transformation and to impart herbicide resistance
XX Disclosure; Page 86-89; 194pp; English.

XX This sequence is the wild type Arabidopsis thaliana threonine
XX dehydratase/deaminase (TD) protein. The invention relates to mutants of
XX the encoded protein, that are feedback insensitive TD mutants. The TD DNA
XX sequence is used as molecular marker (imparting resistance to toxic
XX structural analogues of isoleucine) for selecting transformed cells and

CC to produce transformants with increased levels of isoleucine (and thus
CC better nutritional value) or of intermediates in biosynthesis of
CC isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable
CC biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use
CC of the isoleucine structural analogues as herbicides. The DNA sequences
CC are alternatives for antibiotic resistance markers (which are potentially
CC harmful to the environment). Since no human analog of TD exists (humans
CC can not synthesize isoleucine), it should be safe to use.

XX Sequence 592 AA;

Query Match 7.5% Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;
QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPKDFLSVM-----LEKGSLS-AMRFLT-- 104
Db 52 ppklpplr-----lkvspnslq---ypagylgavpertenaeagsiaameyltnl 99
QY 105 -----AVNLEHP-EMLEKASRELWMRVMSRNEDITE-----POSILAA 141
Db 100 lskvydiaesplqlaklskrigvmlykredlqpvfsfklrgaynmvklpadqlak 159
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQLKETTEAACRYGAFGLPTIVA 188
Db 160 gvicssagnhagqvalsasklgtavimpvttpeikwqavenl----- 203
QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220
Db 204 ---gatvvifgdsydaqahakiraeegltfipp 235

RESULT 15

AA050702
ID AAY05702 standard; Protein; 592 AA.

XX AAY05702;

DT 19-JUL-1999 (first entry)

DE Arabidopsis wild-type threonine dehydratase/deaminase.

KW Threonine dehydratase/deaminase; OMRI gene; feedback inhibition;
KW transgenic plant; selectable marker; isoleucine.
XX Arabidopsis thaliana.

Key Location/Qualifiers
FT Peptide 1..90
FT Protein /note= "transit peptide"
FT Protein 91..592
FT Region /note= "mature protein"
FT Region 486..504
FT Region /note= "regulatory region R4"
FT Region 536..554
FT Region /note= "regulatory region R6"

PN WO9902656-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-US14362.

XX 17-FEB-1998; 98US-0074875.

XX 10-JUL-1997; 97US-0052096.

XX (PURD) PURDUE RES FOUND.

XX Mourad GS;

XX WPI; 1999-120860/10.

XX N-PSDB; AAX25331.

XX

PT New sequences encode mutant threonine dehydratase/deaminase - which
PT is insensitive to feedback inhibition, useful as a selective marker
PT to produce transformed cells resistant to toxic isoleucine analogues
XX Disclosure; Page 50-53; 120pp; English.

XX The present sequence represents Arabidopsis thaliana var. Columbia
CC wild-type threonine dehydratase/deaminase (TD), the first enzyme of
CC the isoleucine biosynthetic pathway. TD is encoded by the OMRI
CC gene (see AAX25331) of A. thaliana. The invention provides nucleotide
CC sequences (see AAX25332-40), originally isolated and cloned from A.
CC thaliana mutated line GM1b (omri/omri), which encode feedback
CC insensitive TD that can be used to transform a wide variety of
CC plants, fungi, bacteria and yeast. The mutated form of TD differs
CC from the wild-type only by 2 point mutations (C to T at nucleotide
CC 1495, and G to A at nucleotide 1631), which result in an R499C amino
CC acid substitution in the regulatory region R4 of TD, and an R544H
CC substitution in the regulatory region R6. These forms of TD are not
CC only insensitive to feedback inhibition by isoleucine, but are also
CC plants and microorganisms which synthesize only wild-type TD.
CC Nucleotide sequences encoding mutated forms of TD can therefore be
CC used to create cells that are insensitive to compounds normally
CC toxic to cells expressing only wild-type TD enzymes, and thus may
CC be used to provide a biochemical selectable marker. Transformants
CC harboring a nucleotide sequence comprising a promoter operably
CC linked to a mutated TD sequence demonstrate increased levels of
CC isoleucine production, and thus provide an improved nutrient source.

XX Sequence 592 AA;

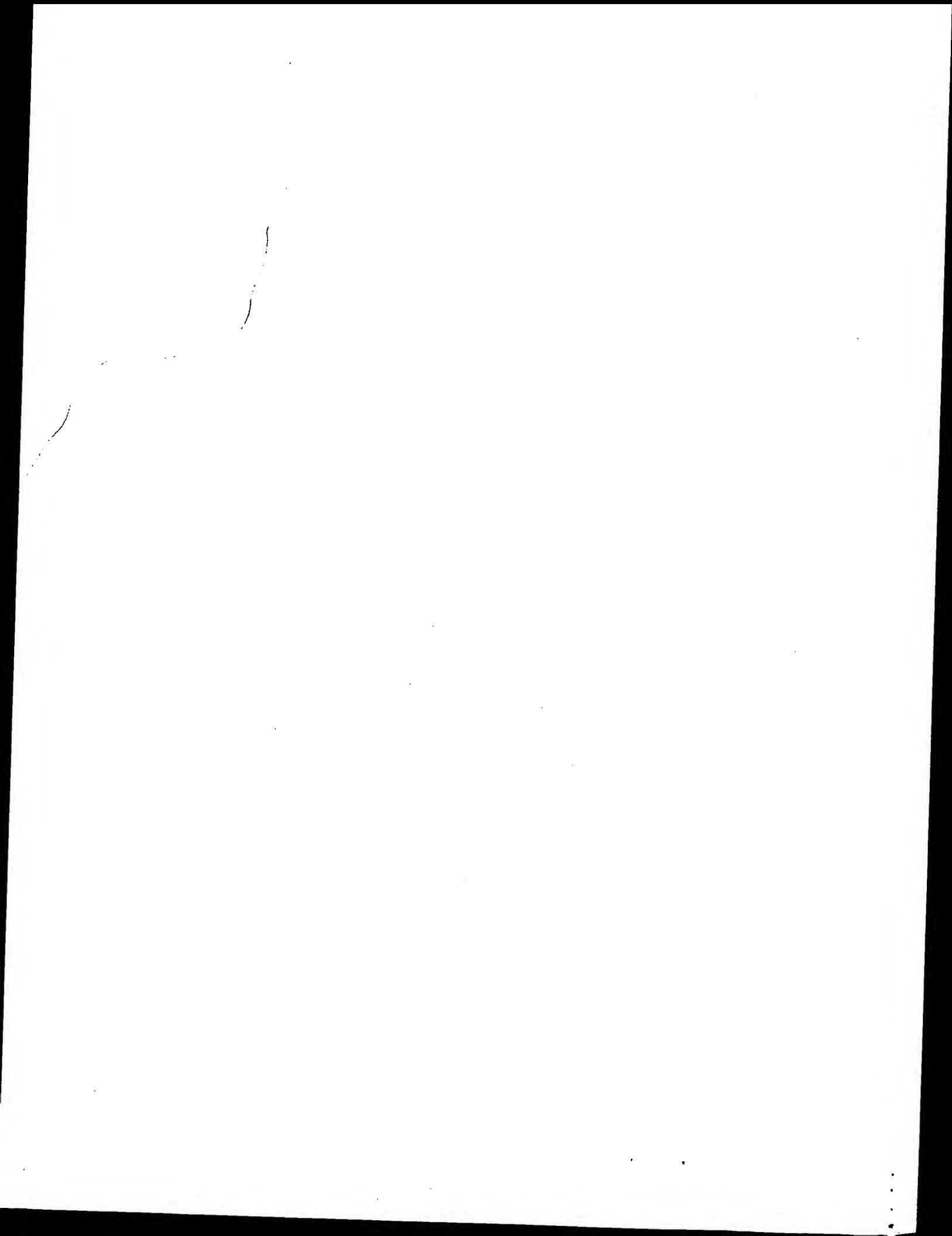
Query Match 7.5% Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPKDFLSVM-----LEKGSLS-AMRFLT-- 104
Db 52 ppklpplr-----lkvspnslq---ypagylgavpertenaeagsiaameyltnl 99
QY 105 -----AVNLEHP-EMLEKASRELWMRVMSRNEDITE-----POSILAA 141
Db 100 lskvydiaesplqlaklskrigvmlykredlqpvfsfklrgaynmvklpadqlak 159
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQLKETTEAACRYGAFGLPTIVA 188
Db 160 gvicssagnhagqvalsasklgtavimpvttpeikwqavenl----- 203
QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220
Db 204 ---gatvvifgdsydaqahakiraeegltfipp 235

Search completed: May 20, 2002, 08:39:12
Job time: 344 sec

us-09-441-723-1.rag

Tue May 21 08:04:00 2002



Sequence 2, Appl1
Sequence 13, Appl1
Sequence 2, Appl1
Sequence 23, Appl1
Sequence 5, Appl1
Sequence 38, Appl1
Sequence 38, Appl1
Sequence 11, Appl1
Sequence 51, Appl1
Sequence 15, Appl1
Sequence 2, Appl1
Sequence 5, Appl1
Sequence 5, Appl1
Sequence 1, Appl1
Sequence 2, Appl1

28 71.5 6.0 543 4 US-09-071-739B-2
29 71.5 6.0 543 4 US-09-181-336-13
30 71.5 6.0 543 4 US-09-260-038B-2
31 71 6.0 587 1 US-07-935-905A-23
32 71 6.0 1090 4 US-09-346-237-5
33 70.5 5.9 615 2 US-08-484-101B-38
34 70.5 5.9 615 4 US-08-714-524D-38
35 70 5.9 401 1 US-08-198-446B-11
36 70 5.9 401 2 US-08-870-693-11
37 69.5 5.9 259 2 US-07-857-224B-51
38 69.5 5.9 529 4 US-08-426-509A-15
39 69.5 5.9 529 5 PCT-US95-05008-15
40 69.5 5.9 1075 1 US-07-623-033-2
41 69 5.8 689 1 US-07-766-351-5
42 69 5.8 689 1 US-08-059-032-5
43 68.5 5.8 689 5 PCT-US91-07290-5
44 68.5 5.8 339 3 US-08-758-280-1
45 68.5 5.8 339 3 US-08-758-280-2

ALIGNMENTS

RESULT 1
US-08-978-174-1
; Sequence 1, Application US/08978174
; Patent No. 6030809
; GENERAL INFORMATION:
; APPLICANT: Shah, Purvi
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ For Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/08/978,174
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0430 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADTUT04
; CLONE: 1554593
US-08-978-174-1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 20, 2002, 08:33:48 ; Search time 13.17 Seconds
(without alignments)
419.148 Million cell updates/sec

Title: US-09-441-723-1
Perfect score: 1185
Sequence: 1 MGPLPRTVEFDVLSYSW.....AHLGKMGPIPPAVNARL 226

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution..

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	3	US-08-978-174-1
2	874	73.8	226	3	US-08-978-174-3
3	80.5	6.8	339	2	US-08-855-714-3
4	76.5	6.5	317	2	US-09-066-075-2
5	76.5	6.5	317	2	US-08-518-615A-2
6	76.5	6.5	317	3	US-08-951-889-2
7	76.5	6.5	317	4	US-09-472-857-2
8	76.5	6.5	497	1	US-08-075-193-4
9	76.5	6.5	497	2	US-08-564-090A-4
10	76.5	6.5	497	5	PCT-US94-06698-4
11	75.5	6.4	1398	1	US-08-750-532-9
12	75.5	6.4	1398	4	US-08-894-818B-8
13	75.5	6.4	1398	4	US-09-445-472-6
14	73.5	6.2	523	2	US-08-473-553A-3
15	73.5	6.2	869	1	US-08-188-582-32
16	73.5	6.2	869	1	US-08-646-715-32
17	73.5	6.2	980	2	US-08-473-553A-6
18	73.5	6.2	985	2	US-08-473-553A-2
19	73	6.2	948	1	US-08-698-551-14
20	73	6.2	948	2	US-08-602-228-14
21	73	6.2	948	2	US-08-533-901B-14
22	73	6.2	948	2	US-08-839-032A-14
23	73	6.2	948	2	US-08-839-031A-14
24	73	6.2	948	5	PCT-US95-12724-14
25	71.5	6.0	461	4	US-09-346-408-8
26	71.5	6.0	532	4	US-09-181-336-15
27	71.5	6.0	543	2	US-08-922-170B-10

Query Match 100.0%; Score 1185; DB 3; Length 226;
 Best Local Similarity 100.0%; Pred. No. 5.3e-126;
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSPSYWLGFELICRYQNIWNNINLQRLPSLTIGIMKDSGNKPPGLLP 60
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 1 MGPLPRTVELFDVLSPSYWLGFELICRYQNIWNNINLQRLPSLTIGIMKDSGNKPPGLLP 60
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 61 RKGLYMANDLKLRLHHLQIPIHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 61 RKGLYMANDLKLRLHHLQIPIHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 121 LMRVWSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNOLKETTEAACRYGA 180
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 121 LMRVWSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNOLKETTEAACRYGA 180
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

RESULT 2
 US-08-978-174-3
 : Sequence 3, Application US/08978174
 : Patent No. 6030809
 : GENERAL INFORMATION:
 : APPLICANT: Shah, Purvi
 : APPLICANT: Hillman, Jennifer L.
 : APPLICANT: Lal, Preeti
 : APPLICANT: Corley, Neil C.
 : TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE
 : NUMBER OF SEQUENCES: 3
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Incyte Pharmaceuticals, Inc.
 : STREET: 3174 Porter Drive
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/978,174
 : FILING DATE: Herewith
 : CLASSIFICATION:
 : PRIOR APPLICATION NUMBER:
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Billings, Lucy J.
 : REGISTRATION NUMBER: 36,749
 : REFERENCE/DOCKET NUMBER: PF-0430 US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 650-855-0555
 : TELEFAX: 650-845-4166
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 226 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : IMMEDIATE SOURCE:
 : LIBRARY: GenBank
 : CLONE: ?
 : US-08-978-174-3

Query Match 73.8%; Score 874; DB 3; Length 226;
 Best Local Similarity 69.5%; Pred. No. 7.8e-91;

Matches 157; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSPSYWLGFELICRYQNIWNNINLQRLPSLTIGIMKDSGNKPPGLLP 60
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 1 MGPAFRLVELFDVLSPSYWLGFELICRYQHLWNKIKLRPALLAGIMKDSGNQPPAMVP 60
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 61 RKGLYMANDLKLRLHHLQIPIHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 61 HKGOVILKEIPLKQLFQVPMSPKDFEGEHVKGKGTYNAMRFLTAVNMEQPEMLEKVSRE 120
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 121 LMRVWSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNOLKETTEAACRYGA 180
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 121 LMRVWSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNOLKETTEAACRYGA 180
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||
 Db 181 FGLPTTVAHVGDQTHMLFGSDRMELLAYLLGKWMGPVPTTLNARL 226
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||

RESULT 3
 US-08-855-714-3
 : Sequence 3, Application US/08855714
 : Patent No. 5939075
 : GENERAL INFORMATION:
 : APPLICANT: Hough, Huo-Shu H.
 : APPLICANT: Warren, Richard L.
 : TITLE OF INVENTION: MUTANTS OF BRUCELLA MELITENSIS
 : NUMBER OF SEQUENCES: 3
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: John Moran, Esq.
 : STREET: HQ USAMRDC, Dept. of Army, Fort Detrick
 : CITY: Frederick
 : STATE: MD
 : COUNTRY: US
 : ZIP: 21702-5012
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/855,714
 : FILING DATE:
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/334,129
 : FILING DATE: 04-NOV-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hendricks, Glenna
 : REGISTRATION NUMBER: 32,535
 : REFERENCE/DOCKET NUMBER: 08/143,692
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (301) 619-2065
 : TELEFAX: (301) 619-7714
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 339 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: both
 : TOPOLOGY: unknown
 : MOLECULE TYPE: NO
 : HYPOTHETICAL: NO
 : ANTI-SENSE: NO
 : ORIGINAL SOURCE:
 : ORGANISM: BRUCELLA MELITENSIS
 : US-08-855-714-3

Query Match 6.8%; Score 80.5; DB 2; Length 339;
 Best Local Similarity 21.7%; Pred. No. 0.77;
 Matches 65; Conservative 25; Mismatches 96; Indels 113; Gaps 14;

QY 3 PLPRTVELFDVLS---PYSWLGFELICRYQNIWNNINLQRLPSLTIGIMKDSGNKPPGLL 59

us-09-441-723-1-rai

Tue May 21 08:04:01 2002

QY 117 ASRE-----LWVRVWSRNE-----ITEPQS----- 137
 Db 103 DPEHKEFRFLWKQIADRYKDYPTLFEILNEPHGNLTPEKNWELLEALKVIRSIDK 162
 QY 138 -----ILAAAEKAGMSAEQAQGLLEKIAIATPKVKVKNQKLETTAEACRYCAFGLPITVAHVDGQ 193
 Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE-----KNSIVTIHYINPFEF----- 203
 QY 194 THMLFGSDRMELLALHLLGCKWMP 217
 Db 204 THQ--GAEWVEGSEKWLGRKWCSP 225

RESULT 5
 US-08-518-615A-2
 ; Sequence 2, Application US/08518615A
 ; Patent No. 5962258
 ; GENERAL INFORMATION:
 ; APPLICANT: Mathur, E., et al.
 ; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/518,615A
 ; FILING DATE: August 23, 1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: FERRARO, GREGORY D.
 ; REGISTRATION NUMBER: 36,134
 ; REFERENCE/DOCKET NUMBER: 331400-20
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-994-1700
 ; TELEFAX: 201-994-1744
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 317 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; STRANDEDNESS:
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN
 ; US-08-518-615A-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;
 Best Local Similarity 21.1%; Pred. No. 2;
 Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;
 QY 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPMELEK 116
 Db 49 HVRIPRWSTHAYAFPPYKIMDRFRKRVDEVINGALKRG-----LAVAINIHVEELMN 102
 QY 117 ASRE-----LWVRVWSRNE-----ITEPQS----- 137
 Db 103 DPEHKEFRFLWKQIADRYKDYPTLFEILNEPHGNLTPEKNWELLEALKVIRSIDK 162
 QY 138 -----ILAAAEKAGMSAEQAQGLLEKIAIATPKVKVKNQKLETTAEACRYCAFGLPITVAHVDGQ 193
 Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE-----KNSIVTIHYINPFEF----- 203

Db 47 PLPTSPMLKWCQSAPPTSLWKRRFLCPRPPIHWK-----SLRTASOKSSFSTKAALK 99
 QY 60 PRK-GLYMANDKLLRHHLIQIPHFQKDFLSVMLEKGL-----SAMRFLTAVNLEH 110
 Db 100 PRPGSWMTKRKSSPRSGALRAWH-PODRRLGVDGKQVRLASLDETQACNAFRAIN-KA 157
 QY 111 PEMLE---KASRELWVRVWSRNEIITEPQSILAAAEKAG----- 146
 Db 158 PAILEGFVEFREV-----SVIAARDRSGNVAIFDLAENVHKDGILATST 202
 QY 147 -----MSAEQAQGLLEKTA-----TPKVKVKNQKLETTAE 173
 Db 203 VPAAISVQTAEAARTAEKILHALDYYGVGLGFEFLVLDKGTLLANEFAPRVHNS-GHWTE 261
 QY 174 ACRYGAF-----GLPI--TVAHVDGQTHMLFGSD-----RMELLAHLGK 213
 Db 262 AACAIQFEQHRAVAGLPLGNTDRHSDCVNENLIGDDIEKVPAILCEKNAVHLHYGKK 320

RESULT 4
 US-09-066-075-2
 ; Sequence 2, Application US/09066075
 ; Patent No. 5925749
 ; GENERAL INFORMATION:
 ; APPLICANT: Mathur, E., et al.
 ; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066,075
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/518,615
 ; FILING DATE: August 23, 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: FERRARO, GREGORY D.
 ; REGISTRATION NUMBER: 36,134
 ; REFERENCE/DOCKET NUMBER: 331400-20
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-994-1700
 ; TELEFAX: 201-994-1744
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 317 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; STRANDEDNESS:
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN
 ; US-09-066-075-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;
 Best Local Similarity 21.1%; Pred. No. 2;
 Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;
 QY 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPMELEK 116
 Db 49 HVRIPRWSTHAYAFPPYKIMDRFRKRVDEVINGALKRG-----LAVAINIHVEELMN 102

QY 194 THMLFGSDRMELLALLGCKWMP 217
||| : : : ||| : : :
Db 204 THQ--GAEWEGSEKWLGRKWGP 225

RESULT 6
US-08-951-889-2
; Sequence 2, Application US/08951889
; Patent No. 6008032
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; TITLE OF INVENTION: Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-951-889-2

Query Match 6.5%; Score 76.5; DB 3; Length 317;
Best Local Similarity 21.1%; Pred. No. 2;
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;
QY 76 HLQPIH-----FP-----KDFLSVMLEKSGLSAMRFLTAVNLEHPMLEK 116
||| : : : ||| : : :
Db 49 HVRIPIRSTHAYAPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHVEELMN 102
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS----- 137
Db 103 DPEHKERFLALWKQIADRYKDPETLFFELNPNHGNLTPEKNELLEALKVIRSIDK 162
QY 138 -----ILAAEKAGSAEQAGLLEKIATPKVNQLKETTAAACRYGAFGLPIIVAHVDGQ 193
Db 163 KHTIIIGTAEMGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEP----- 203
QY 194 THMLFGSDRMELLALLGCKWMP 217
||| : : : ||| : : :
Db 204 THQ--GAEWEGSEKWLGRKWGP 225

RESULT 8
US-08-075-193-4
; Sequence 4, Application US/08075193
; Patent No. 5547868

RESULT 7
US-09-472-857-2
; Sequence 2, Application US/09472857
; Patent No. 6245547
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; TITLE OF INVENTION: Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/472,857
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/951,889
; FILING DATE:
; APPLICATION NUMBER: 08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-472-857-2

Query Match 6.5%; Score 76.5; DB 4; Length 317;
Best Local Similarity 21.1%; Pred. No. 2;
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;
QY 76 HLQPIH-----FP-----KDFLSVMLEKSGLSAMRFLTAVNLEHPMLEK 116
||| : : : ||| : : :
Db 49 HVRIPIRSTHAYAPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHVEELMN 102
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS----- 137
Db 103 DPEHKERFLALWKQIADRYKDPETLFFELNPNHGNLTPEKNELLEALKVIRSIDK 162
QY 138 -----ILAAEKAGSAEQAGLLEKIATPKVNQLKETTAAACRYGAFGLPIIVAHVDGQ 193
Db 163 KHTIIIGTAEMGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEP----- 203
QY 194 THMLFGSDRMELLALLGCKWMP 217
||| : : : ||| : : :
Db 204 THQ--GAEWEGSEKWLGRKWGP 225

RESULT 8
US-08-075-193-4
; Sequence 4, Application US/08075193
; Patent No. 5547868

GENERAL INFORMATION:
APPLICANT: MILLER, WALTER L.
APPLICANT: HARIKRISHNA, JENNIFER A.
APPLICANT: BLACK, STEPHEN M.
TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306

* COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

- CURRENT APPLICATION DATA: US/08/075,193
APPLICATION NUMBER: US/08/075,193
FILING DATE: 09-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: NEELEY Ph.D., RICHARD L.
REGISTRATION NUMBER: 30,092
REFERENCE/DOCKET NUMBER: UCAL-236/000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-075-193-4

Query Match 6.5%; Score 76.5; DB 1; Length 497;
Best Local Similarity 25.4%; Pred. No. 3.9;
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;

QY 14 VLSPYSWLGFELCRYONIWINLQLPSLITGIMKDSGNKPPGCLLPKRG-LYWANDLKL 72
DB 195 LLTPPEHLEALLLCQRTDITKAAL-----GVLRQSRVKTVMWVGRRGPLOVAFIKE 246
QY 73 LRHHLOI---PIHFPRKDFL-----SVMLEKGSLSAMRFLTAVNLEHPMLEKASREL 121
DB 247 LREMIOQLPGARPILDPVDFLGLQDKIKEVPRPKRLTELLRTAT--EKPGPAEAARQAS 304
QY 122 WMRVWSNRNEDITEPQSILAAAKAGMAEQAGLLEKIATPKVKNQLKETTEA 174
DB 305 ASRAWGL-RFFRSQQVLPSPD-----GRRAGV--RLAVIRLEG-VDEATFA 348

RESULT 9
US-08-564-090A-4
; Sequence 4, Application US/08564090A
; Patent No. 5939318
; Patent No. 5939318 5741703
; GENERAL INFORMATION:
; APPLICANT: MILLER, WALTER L.
; APPLICANT: HARIKRISHNA, JENNIFER A.
; APPLICANT: BLACK, STEPHEN M.
; TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD LLP
; STREET: FIVE PALO ALTO SQUARE
; STREET: 3000 EL CAMINO REAL
; CITY: PALO ALTO

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 213-977-1001
 TELEFAX: 213-977-1003
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-06698-4

Query Match 6.5%; Score 76.5; DB 5; Length 497;
 Best Local Similarity 25.4%; Pred. No. 3.9;
 Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;
 QY 14 VLSPYSWGLFELCRQNTWNLQLRPSLITGTMKDSGNKPPGLPRKG-LYMANDLKL 72
 Db 195 LITPPEHLEALLCQRTDITKAAL-----GVLROSRTKTVWLVRGRGLQVAFITKE 246
 QY 73 LRHHLOI-----PIHPKDFL-----SVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121
 Db 247 LREMIOLPGARILPDVDFLQDKIKEVPRPRKRLTELLRTAT--EKPGPAEAARQAS 304
 QY 122 WMVYWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKLEKIATPKVKNQKLETTA 174
 Db 305 ASRAWGL-REFRSPQVLFSPD-----GRRAGV--RLAVTRLEG-YDEATRA 348

RESULT 11
 US-08-750-532-9
 ; Sequence 9, Application US/08750532
 ; Patent No. 5756339
 ; GENERAL INFORMATION:
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
 ; STREET: 419 Seventh Street N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/750,532
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP95/01095
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 1994/130236
 ; FILING DATE: 13-JUN-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 1994/173912
 ; FILING DATE: 26-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROWDY, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: MITTA-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1398 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-750-532-9

Query Match 6.4%; Score 75.5; DB 1; Length 1398;
 Best Local Similarity 16.8%; Pred. No. 24;
 Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;
 QY 50 DSGNKPGLLPKGLY--MANDLKLRLHHLOI----- 79
 Db 1097 DKADFAVGLTPAEGVLGEARNYTLIVKHALTLPEVPNATVIIGNVTYLTDEGNTVTYTA 1156
 QY 80 PIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRELWVRVWNRNEDITEPOSIL 139
 Db 1157 PTKLGSDETTIVIVKKNFNTLEKTFQITVSEPEITE-----EDINEPKLAM 1202
 QY 140 AAEKAG--MSAE-QAQLLEKIAT-----PKVKNQKLETTAAACR 177
 Db 1203 SSPEANATIVSMESEGGVKKTVTVEITINGTANETATIVVPVKKAENIEVSDHVIS 1262
 QY 178 YG-----AFGLPITVAHVDTGTHMLFGSDRMELLA-HLIGEKW 214
 Db 1263 YSIEGEYAKVVIITVKFASPVTVT---VTYIYAGPRVSIILTNFLGYSW 1310

RESULT 12
 US-08-894-818B-8
 ; Sequence 8, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA-1
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-8

Query Match 6.4%; Score 75.5; DB 4; Length 1398;
Best Local Similarity 16.8%; Pred. No. 24;
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;
QY 50 DSGNKPGLPRKGLY--MANDLKLRLHQLI-----79
Db 1097 DKADFAVGLTPAEGVLGEARNYTLIVKHALTLEPVPNATVIIGNYTLTDENGTVFTYA 1156
QY 80 PIHPKDFLSVMEKSGLSAMRFLTAVNLEHPMELEKASRELWMRVWSNEDITEPQSIL 139
Db 1157 PTKGSDIIVVKKENFTLEKTFQITVSEPEITE-----EDINEPKLAM 1202
QY 140 AAARKAG--MSAE-QAQLLEKIAT-----PKVNQLKETTEAACR 177
Db 1203 SSPEANATIVSMESEGGVKKTVTVEITINGTANETATIVVPVPPKAENIEVSGDHVIS 1262
QY 178 YG-----AFGLPITVAHVDGQTHMFLGSDRMELLA--HLLGEKW 214
Db 1263 YSIEGEYAKYVIITVKFASPVTVT-----VTYIVAGPRVSILTLNFLGYSW 1310

RESULT 13
US-09-445-472-6
Sequence 6, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: SHIMOJO, TOMOKO
APPLICANT: ASADA, KIYOZO
APPLICANT: KATO, IKUNOSHIN
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472
CURRENT FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 6
LENGTH: 1398
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-445-472-6

Query Match 6.4%; Score 75.5; DB 4; Length 1398;
Best Local Similarity 16.8%; Pred. No. 24;
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;
QY 50 DSGNKPGLPRKGLY--MANDLKLRLHQLI-----79
Db 1097 DKADFAVGLTPAEGVLGEARNYTLIVKHALTLEPVPNATVIIGNYTLTDENGTVFTYA 1156
QY 80 PIHPKDFLSVMEKSGLSAMRFLTAVNLEHPMELEKASRELWMRVWSNEDITEPQSIL 139
Db 1157 PTKGSDIIVVKKENFTLEKTFQITVSEPEITE-----EDINEPKLAM 1202
QY 140 AAARKAG--MSAE-QAQLLEKIAT-----PKVNQLKETTEAACR 177
Db 1203 SSPEANATIVSMESEGGVKKTVTVEITINGTANETATIVVPVPPKAENIEVSGDHVIS 1262

QY 178 YG-----AFGLPITVAHVDGQTHMFLGSDRMELLA--HLLGEKW 214
Db 1263 YSIEGEYAKYVIITVKFASPVTVT-----VTYIVAGPRVSILTLNFLGYSW 1310
RESULT 14
US-08-473-553A-3
Sequence 3, Application US/08473553A
Patent No. 5859338
GENERAL INFORMATION:
APPLICANT: Meyerowitz, Elliot M.
APPLICANT: Clark, Steven E.
APPLICANT: Williams, Robert W.
TITLE OF INVENTION: Plant Clavatal Nucleic Acids,
TITLE OF INVENTION: Transformed Plants, and Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,553A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 523 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-473-553A-3

Query Match 6.2%; Score 73.5; DB 2; Length 523;
Best Local Similarity 21.8%; Pred. No. 9.2;
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;
QY 36 NLQRLPSLITGIMKDSGNKPP---GLLPKGLYMANDLKLRLHQLIPIHPKDFLSVML 92
Db 194 NLKHLHLFLHNNLTGHIPPELSGLVSLKSL---DLISI---NOLTGTEPOSFINL-- 243
QY 93 EKSLSAMRFLTAVNL-----EHPMELEKASRELWMRVWSNEDITEPQSILAAAE-- 143
Db 244 --GN-----ITLINLFRNNLYGQIPEAIGELPKLEVFYEWNNFTLQLPANLGRGNLI 295
QY 144 KAGMSAEQAQGLL-----EKIATPK-----VKNQLKET 171
Db 296 KLDVSDNHLTGILPKDLCKGEKLEMLLSNNFFGPTPEELGKCKSLTKIRIVKNLLNGT 355
QY 172 TEACRYGAFGLP-ITVAHV-----DQOTHMFLGSDRMELLAHLLGEKW--GP1PPAV 222
Db 356 VPA---GLFNLPLVTIIELTDNFFSGELPVTMTSGVDLDOI---YLSNNWFSGEIPPAI 407
RESULT 15

US-08-188-582-32
; Sequence 32, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoaki
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 869 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-582-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;
Best Local Similarity 20.5%; Pred. No. 20;
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;

QY	32	IWNINQLRPSLITGIMKDSGNKPPGLLPKGLYMANDLKLRRHHQIPIHFPPKDFLSVM	91
Db	318	VWKEGKQMPQLLQAMQVEKGATGTSLSPP---HLPGLAICSRSGAVCLWSPEDGLRQI	373
QY	92	LEKGLSAMRELTAVNLEHPEMPELEKASRELWMRVWSRNNEDITEPQSTILAAAEKAGMSAQ	151
Db	374	-----YRDPETLVFRDSSWR--WA---DFTAHPRLVTGDRGTGVKMLD	412
QY	152	AQG-----LLEKATPKVKKQKETTAAACRYGA-----FG-----LPITVAHV	192
Db	413	TQPPPGGGLLLFRLG-----AEASCQKGERVLLTQYLGHSSPKCLPPTLHLVCT	461
QY	193	QTHMFGSDRMLLAHLGKWMGPPI-PAVNARL	226
Db	462	QFSILVLDRLPLVPMPL---KWNHGLPSPLLLARL	493

Search completed: May 20, 2002, 08:39:32
Job time: 344 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:34:43 ; Search time 19.29 Seconds
(without alignments)
1125.774 Million cell updates/sec

Title: US-09-441-723-1
Perfect score: 1185
Sequence: 1 MGPLPRTVELFDVLSYPSW.....AHLLEGKMGPIPAVNARL 226

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.5	27.7	226	2 T27747	hypothetical prote
2	298.5	25.2	225	2 T34201	hypothetical prote
3	198	16.7	195	2 G85629	hypothetical prote
4	164.5	13.9	194	2 S72164	2-hydroxychromene-
5	147.5	12.4	206	2 H87328	conserved hypotet
6	145.5	12.3	312	2 B88492	protein T07E3.3 li
7	140	11.8	197	2 T31286	2-nitrotoluene dio
8	125	10.5	33	2 S17164	glutathione transf
9	124	10.5	197	2 AB3212	2-hydroxychromene-
10	114	9.6	253	2 G87394	hypothetical prote
11	95	8.0	203	2 C55552	2-hydroxychromene-
12	91	7.7	581	2 H72425	ABC transporter, A
13	90.5	7.6	410	2 C69127	flavoprotein A hom
14	90.5	7.6	600	2 D95899	probable dipeptid
15	89.5	7.6	962	1 SNECPI	pitriylsin (EC 3.4
16	89.5	7.6	962	2 F91088	proteinase III [im
17	89.5	7.6	962	2 H85933	proteinase III [im
18	89	7.5	199	2 I49343	probable isomerase
19	89	7.5	592	2 T51712	threonine dehydrat
20	88.5	7.5	613	2 A35296	secretogranin II p
21	87.5	7.4	962	2 AC0865	protease III precu
22	87	7.3	926	2 E83375	probable glycosyl
23	85.5	7.2	224	2 A13595	frne protein (impo
24	84.5	7.1	316	2 B71301	probable tRNA delt
25	84.5	7.1	971	2 S45595	probable membrane
26	83.5	7.0	488	2 A40367	catalase (EC 1.11.
27	83.5	7.0	700	1 S09748	kinesin-related pr
28	83.5	7.0	1007	2 C84668	probable receptor-
29	82.5	7.0	263	2 T51169	hypothetical prote

frne protein VCA01
polyketide biosynt
secretogranin II -
DNA-directed DNA p
serine/threonine k
hypothetical prote
protein kinase CDC
hypothetical prote
topoisomerase I -
catalase (EC 1.11.
heat-stable entero
rRNA methylase Spo
ABC transporter (A
hypothetical prote
conserved hypotet
alanine dehydrogen

ALIGNMENTS

RESULT 1
T27747
hypothetical protein ZK1320.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27747
R:Berks, M.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z20414
A:Accession: T27747
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-226 <WILL>
A:Cross-references: EMBL:Z46934; PIDN:CAA87039.1; GSPDB:GN00020; CESP:ZK1320.1
A:Experimental source: Clone ZK1320
C:Genetics:
A:Gene: CESP:ZK1320.1
A:Map position: 2
A:Introns: 23/3; 177/3

Query Match 27.7%; Score 328.5; DB 2; Length 226;
Best Local Similarity 33.0%; Pred. No. 7.2e-21;
Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;
QY 1 MGPLPRTVELFDVLSYPSWLGFEILCRYQNIW-NINLQLRPSLITGIMKDSGNKPPGLL 59
Db 1 MPKLER-IDFYEDVISFYSAIEVFQKLETKQWGVITIRYIPFFLGAVMKESGNRRPPAML 59
QY 60 PRKGLYMANDLKLRLHLLQIPHFPPKDFLSVNLKSGLSAMRFLTAVNLEHPMLEKASR 119
Db 60 PARSIMMTDLKRTAKFWIDPLTPPPLFMEMWIKRYTTTGAMKVLVLVLSQDKELMLRAAR 119
QY 120 ELWMRVWSRNEDITEPQSIILAAAEKAGMSAQGLLEKIATPKVNQLKTKETEAACRYG 179
Db 120 EMVRLWSSEKIEFDQDFVEVLKAVGV--KNPEQIVKSKDEKIKILMENTNKGVLDLM 177
QY 180 AFGLP-ITVAHVQDQTHMFGSDRMELLAHLLEKMGPIIP 219
Db 178 AVGAPWINVTEGSEHSFFGSDRFLIADLLQQP--OPLP 216

RESULT 2
T34201
hypothetical protein D2024.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34201
R:Du, Z.; Gattung, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid D2024.
A:Reference number: Z21488

A:Accession: T34201
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-225 <DU2>
A:Cross-references: EMBL:U41011; PIDN:AAA82289.1; CESP:D2024.7
C:Genetics:
A:Gene: CESP:D2024.7
A:Introns: 51/1; 177/3

Query Match 25.2%; Score 298.5; DB 2; Length 225;
Best Local Similarity 33.0%; Pred. No. 2.7e-18;
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;
QY 3 PLPRTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGL---L 59
Db 2 PNKVVKKFFDWSYSGFEGITRHSVWKPIQMKPFFAGVVRITEN--PGLPLRI 59
QY 60 PRKGLYMANDLKLRLHHLQIPIHFPKDLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119
Db 60 PIKERYMHKDLFLSAQYWGIPFLPKDYTNMLNTSSIVPQILVASQLRONVLMEDVAR 119
QY 120 ELWMRVWSRNEI-TEPQSLAAAEKAGSAEQAGLLEKATPKVKNLKETTAACRY 178
Db 120 GLWHRFYAGKPIFTKSKQ---VAEVLRLDLHVKDVELVMMSDSAEVKNILRENTDEAIGN 176
QY 179 GAFGLP---ITVAHVDGQT-HMLFGSDRMELLALHLLGKWMGPI 218
Db 177 GCFGAPWHITDGH--GKVLQIVFGSDRLPQVADFLAEPKGP 218

RESULT 3
G83629
hypothetical protein PA0118 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83629
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lm,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: G83629
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <STO>
A:Cross-references: GB:AE004450; GB:AE004091; NID:g9945943; PIDN:AA03508.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0118
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

*Query Match 16.7%; Score 198; DB 2; Length 195;
Best Local Similarity 26.1%; Pred. No. 9.9e-10;
Matches 55; Conservative 41; Mismatches 93; Indels 22; Gaps 5;
QY 4 LPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPRKG 63
Db 1 MSKQIEFFDFGSPPTTAWLQPLPRAAAGASIAWRPMLLGGVFKATGNHSPIEVPAKG 60
QY 64 LYMANDLKLRLHHLQIPI---HPPKDFLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119
Db 61 RYTHDLARVAKRYGVPLAFNPAPINTLTLM-----RGAQGYLGG-----EGFQYFLK 109
QY 120 ELWMRVWSRNEI-TEPQSLAAAEKAGSAEQAGLLEKATPKVKNLKETTAACRYG 179
Db 110 AVEALWVRQNLGKPEVAQVLAEGADPDDE---FLRLVGDEQVKEGLKATTEEAARRG 166
QY 180 AFGCLPITVAHVDGQTHMLFGSDRMELLALHL 210
Db 167 VFGAPSFV---GDQLFFGQDRDLDFVAEVL 193

RESULT 4

S72164
2-hydroxychromene-2-carboxylate isomerase homolog - Rhizobium leguminosarum bv.
C:Species: Rhizobium leguminosarum bv. viciae
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 29-Sep-1999
C:Accession: S72164
R:Brito, B.; Palacios, J.M.; Ruiz-Argueso, T.; Imperial, J.
Biochim. Biophys. Acta 1308, 7-11, 1996
A:Title: Identification of a gene for a chemoreceptor of the methyl-accepting ty
A:Reference number: S72162; MUID:96328256
A:Accession: S72164
A:Molecule type: DNA
A:Residues: 1-194 <BRI>
A:Cross-references: EMBL:U23040; NID:g780654; PIDN:AA044312.1; PID:g780657
A:Experimental source: strain UPW791
C:Genetics:
A:Gene: plasmid
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 13.9%; Score 164.5; DB 2; Length 194;
Best Local Similarity 25.2%; Pred. No. 7.4e-07;
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;
QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64
Db 2 RTLDFYDYRSPYSFLALSQVRK---MDVEIAFHPLIEIGDLMKQGVNPTTITCAPKGR 57
QY 65 YMANDLKLRLHHLQIPIHFPKDFLSV-----MLEKSGLSAMRFLTAVNLEHPEMLEK 116
Db 58 YVMTDIQRAVHYGVSLSNHPHQLLEIDASRLRLATLVAGQLGAMP--TAV----- 105
QY 117 ASRELMMRVWSRNEI-TEPQSLAAAEKAGSAEQAGLLEKATPKVKNLKETTAAAC 176
Db 106 --EAFINAIWSAPAPLATAAAEVAALVGAAGLDAEE---LAERMDPEAAQDLDEATANAV 160
QY 177 RYGAFLPITVAHVDGQTHMLFGSDRMELL-AHL 209
Db 161 SRGVFGAPTLFV---GDEMFFGNDRLHFMQGH 190

RESULT 5

H87328
conserved hypothetical protein CC0643 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H87328
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidell
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Frae
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87328
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AE005673; NID:gl3421856; PIDN:AAK22628.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0643
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 12.4%; Score 147.5; DB 2; Length 206;
Best Local Similarity 24.0%; Pred. No. 2.3e-05;
Matches 50; Conservative 43; Mismatches 86; Indels 29; Gaps 8;
QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPP-----GLLP 60
Db 8 KSIDFIDFGSPNAYLSWKLLPEATAAGAAVRLIPCLLGGIFKATGNQAPMTAFGGVK 67

Qy 61 RKGLYMANDLKLRRH-...LQIPHPKDFLSVLMLEKSGLSAMRELTAVNLEHPMELEK 116
 Db 68 KMDYEMLETRFRFAHGLAFRENPFPVN--TLLMRGMIAAQRIGVA-----E 115
 Qy 117 ASRELMMR-VMSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKATPKVKNOLKETTAA 175
 Db 116 AYLEAMLKMGWEDGLKDDPEVEVATANAAGL---DGAALLAATGDAEVKAEVLVANTEAA 172
 Qy 176 CRYGAFGLPITVAHVGDGTHMLFGSDRM 203
 Db 173 VARGTEGIPTFFV---GE-EIFFGKERL 196

RESULT 6
 B8492
 protein T07E3.3 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: B88492
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: B88492
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-312 <STO>
 A:Cross-references: GB:chr_III; PIDN:AAA21082.1; PID:g532470; GSPDB:GN00021; CESP:T07E3.
 C:Genetics:
 A:Gene: T07E3.3
 A:Map position: 3

Query Match 12.3%; Score 145.5; DB 2; Length 312;
 Best Local Similarity 22.6%; Pred. No. 5.9e-05;
 Matches 51; Conservative 46; Mismatches 100; Indels 29; Gaps 6;

Qy 1 MGPLRTVELFVDVLSVPSY-...-LGFEILCRYQ-----NWNINLQ 38
 Db 1 MAPLR-VKCYEDVVCNPNWITIALTSNLSLPERIDFEPVCDKIGILHNAQIWNORRQ 59
 Qy 39 LRPSLI--TGIMKDSNGKPPGLPRGLYMAND--LKLRRHHLQIPHPKDFLSVMLE 93
 Db 60 VHSRLWTKIEVPEQAEESEETSEMGILQIDRGRKLLGCERVPVVDKNTKYTAVA 119
 Qy 94 KGLSAMRELTAVNLEHPMELEKASRELMMRVMSRNEDITEPQSILAAAEKAGMSAEQAQ 153
 Db 120 KGSVLPQLFTSIREQYPDLYERAIHHLGRKRLWEQRLPVHYGCHMSTVCRELGISFKIAE 179

Qy 154 GLEKATPKVKNOLKETTAACRYGAFGLPITVAHVD--GOTHMLF 198
 Db 180 DIVARLSSPNRSILHKNCKEAVDFKLTEAPGLILLTDEGDTIKIF 225

RESULT 7
 T31286
 2-nitrotoluene dioxygenase (EC 1.14.-.-) Rieske iron-sulfur component - Sphingomonas arc
 C:Species: Sphingomonas aromaticivorans
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T31286
 R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; G
 submitted to the EMBL Data Library, July 1998
 A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati
 A:Reference number: Z20992
 A:Accession: T31286
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-197 <ROM>
 A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378427; PIDN:AAD04010.1
 C:Genetics:
 A:Gene: nahd

A:Genome: plasmid pNLI
 C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase
 C:Keywords: oxidoreductase

Query Match 11.8%; Score 140; DB 2; Length 197;
 Best Local Similarity 23.1%; Pred. No. 9.7e-05;
 Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;

Qy 4 LPRTVELFVDVLSVPSYWLGFELLCRYQNTWNNINLQRLPSLITGIMKDSGNKPPG---LLP 60
 Db 1 MTRTIDFYDFISPFYLAQLKLPETARAAGCTVDWPDIDPEAKIAAGNYGPNREVLP 60
 Qy 61 RKGLYMANDLKLRRHHLQIPHPKDF-----LSVMLEKSGLSAMRELTAVNLEHPML 114
 Db 61 -KIKVMKADLERWAERYGVPLTFPASACADWNCVLFAREHGKAEAFV-----109
 Qy 115 EKASRELMMRVMSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKATPKVKNOLKETTAA 174
 Db 110 -----DAYRRINGQIDPGDRNELAAACATAAGL---DPAALIAVESPAGQNEYRKARSQ 161

Qy 175 ACRYGAFGLPITVAHVGDGTHMLFGSDRMELLAHL 210
 Db 162 AIQRGVYGAFL--MFVDDQ--IPWGNDRDLFLAEYL 193

RESULT 8
 SL7164
 glutathione transferase (EC 2.5.1.18) 13 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Nov-1993 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
 C:Accession: SL7164
 R:Harris, J.M.; Meyer, D.J.; Coles, B.; Ketterer, B.
 Biochem. J. 278, 137-141, 1991
 A:Title: A novel glutathione transferase (13-13) isolated from the matrix of rat
 A:Reference number: SL7164; MUID:91354194
 A:Accession: SL7164
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-33 <HAR>
 C:Keywords: transferase

Query Match 10.5%; Score 125; DB 2; Length 33;
 Best Local Similarity 71.9%; Pred. No. 0.00018;
 Matches 23; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PLPRTVELFVDVLSVPSYWLGFELLCRYQNIWN 34
 Db 2 PAPRVLEFVDVLSVPSYXGLFEVLXRYOHLXN 33

RESULT 9
 AB3212
 2-hydroxychromene-2-carboxylate isomerase [imported] - Agrobacterium tumefaciens
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AB3212
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.;
 : Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gord
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C5
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AB3212
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-197 <KUR>
 A:Cross-references: GB:AE008687; PIDN:AAL46112.1; PID:g17743877; GSPDB:GN00188
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:

RESULT 11
3552
hydroxychromene-2-carboxylate isomerase - *Pseudomonas putida* plasmid NAH7
Species: *Pseudomonas putida*
Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 29-Sep-1999
Accession: C35552

```

Query Match      7.7%; Score 91; DB 2; Length 581;
Best Local Similarity 19.0%; Pred. No. 6.4;
Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 6;

      y      28  RYQNIW-----NINLQRPSLITGIMKDSGNKPPGL-----LPRKGLYMAN 68
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
      b      343 RFENWFSYDGKNWVLKDNILDFQPKLYAIVGETGGCKSTLMSLINGLYIFQKGNIFD 402
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

      y      69  DLKLLRHHLQIP-----IHFPKDFL-----SVMLEKSGLSAMRFLTAVNLEHP-EMLE 115
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
      b      403 EIPLEYNLKLVRKQIAAVQDVLVLESGTILNDIRLFDSEIPERVLSEALKRVHADIIE 462
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

```

```
QY 116 KASRELWVRWNRNEDITEPOSTILAAAEKAGM-----SAEQAGLLEKIATPKVKNLK 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 RUPGGYIEVERGTTLSAGERQLIALARAVLFDKIFILDEATSNVDVITETRIQEALE 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 ETE 173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 ELSK 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
C69127
flavoprotein A homolog (II) - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C:Accession: C69127
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Oiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.;
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: C69127
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <MTH>
A:Cross-references: GB:AE000809; GB:AE000666; NID:G2621265; PIDN:AAB84726.1; PID:G262126
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: WPH220
A:Start codon: TTG
C:Superfamily: Methanobacterium flavoprotein A
C:Keywords: flavoprotein

      Query Match 7.6%; Score 90.5; DB 2; Length 410;
      Best Local Similarity 25.1%; Pred. No. 4.5;
      Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;

QY 13 DVLSPSYWLGFELC-----RYQNIWINLQRLPSLTGIMKDSGNK----- 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 EVLKRY---GSEIICTAKAAAGLROHYSIPQDTPMQ---TVKGDSDTLGGKTLTFLEAP 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 ---PPG---LLPRKGLYMANDLKLRHLLQIPHPKDFLSVMLEKGSLSAMRF-----L 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 MLHPDMSFTLLEEGILFSND--AFQHLICISKRPDKDQPEAVLMD--AAKMFYANLL 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 TAVNLEHPMELEKAS--RELWM-----RVWSRNEDITEPOSILAAAEKAGMSA 149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 TPLS---PLVLRKFSEVKELGLEKIGMTAPSHGQIW-----TEPLKIITAA-----YT 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 EQAQLLEKIATPKVKNLKETEACRYGAGFLPITVAHVDGTHMFLGSDRMELLAHL 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 DWATCKCRDKAT-IYDTHYSTRMLAHAAEGL--MAADVDSMHFLHEDERSEIVKNI 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 LGEK-----WMGPPI 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 LESKAVFIGSPTFNGPFP 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
D95899
probable dipeptide ABC transporter permease and ATP-binding protein SMB20478 [imported]
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95899
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95899
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <KUR>
```

```
A:Cross-references: GB:AL591985; PIDN:CAC48860.1; PID:gl5140333; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-H
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher,
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Le
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Y
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMB20478
A:Genome: plasmid

      Query Match 7.6%; Score 90.5; DB 2; Length 600;
      Best Local Similarity 21.6%; Pred. No. 7.4;
      Matches 33; Conservative 27; Mismatches 44; Indels 49; Gaps 6;

QY 34 NINLQRLPSLTGIMKDSGN-----KPPGLPRKGLYMANDLKLRHLLQI 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 DVSLHVKFGCECLGILGESGSKSVTALSVMGLVASPPGVIRNGAVILCND----- 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 PIHFPKDFLS-----VMLEKGSLSAMRFITAVNLEHPMELEKASRELWVRWNRNEDITE 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 -----DVLSPMPETRLIAKRGSLAYVFDPLTLTHP--MYPVGQV-----EEATAA 435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 POSILAAAEKAGMSAEQAQGLLEKIATPKVKNQ 167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 HQQVRAAERR-----EKAVALLEKVGIPDARER 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
SNECPI
pitriylsin (EC 3.4.24.55) precursor [invalidated] - Escherichia coli
A:Alternate names: endopeptidase Pi; proteinase III
C:Species: Escherichia coli
C:Date: 31-Mar-1993 #sequence_revision 31-Oct-1997 #text_change 18-Feb-2000
C:Accession: F65064; A29093; A25765; B25532
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.
Gene 54, 185-195, 1987
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Esc
A:Reference number: A29093; MUID:88005781
A:Accession: A29093
A:Molecule type: DNA
A:Residues: 1-962 <BLAT>
A:Cross-references: GB:AE000365; GB:U00096; NID:G2367163; PIDN:AAC75860.1; PID:G
A:Experimental source: strain K-12, substrain MG1655
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.
Gene 54, 185-195, 1987
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Esc
A:Reference number: A29093; MUID:88005781
A:Accession: A29093
A:Molecule type: DNA
A:Residues: 1-276; 'HYHSLR' 283, 'W' 285-296 <CLA>
A:Cross-references: GB:M17095; NID:gl47390; PIDN:AAA24436.1; PID:gl47391
A:Experimental source: strain K12
A:Note: part of this sequence, including the amino end of the mature protein, wa
R:Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.
Nucleic Acids Res. 14, 7695-7703, 1986
A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding I
A:Reference number: A25765; MUID:87040734
A:Accession: A25765
A:Molecule type: DNA
A:Residues: 1-962 <FIN>
A:Cross-references: GB:X06227; NID:q42560; PIDN:CAA29576.1; PID:q42561
R:Becker, A.B.; Roth, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992
A:Title: An unusual active site identified in a family of zinc metalloendopeptid
A:Reference number: A38854; MUID:922337263
```

A:Contents: annotation; active site
C:Genetics:
A:Gene: ptr
A:Map position: 61
C:Function:
A:Description: endopeptidase degrades small peptides [validated, MUID:92237263]
A:Pathway: protein degradation
C:Superfamily: insulysin
C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein degradation
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-962/Product: ptilysin #status experimental <MAT>
F:88,92/Binding site: zinc (His) #status experimental
F:91/Active site: Glu #status experimental

Query Match 7.6%; Score 89.5; DB 1; Length 962;
Best Local Similarity 26.3%; Pred. No. 17;
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;
QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----OQLLEKIATPKVKNQ 167
Db 731 WCRNKDVVDKQSVIFEKAGNSTDSALAAVFVPTGYDEYTSAYSSLLGOIVOPWFYNQ 790
QY 168 LKETTEAACRYGAFGLPTTVAHVDOGQTHMFGSDRMELLALLGEXMGPIPPA 221
Db 791 LR--TEEQLCYAVFAFPMSVGRQWGMGFLQSNDKQ---PSFLWERYKAPFFPTA 839

Search completed: May 20, 2002, 08:39:58
Job time: 315 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:39:33 ; Search time 26.65 Seconds
(without alignments)
1467.049 Million cell updates/sec

Title: US-09-441-723-1
Perfect score: 1185
Sequence: 1 MGPLPRTVELFDVLSPSVW.....AHLLEKWMGPIPPAVNARL 226

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL19:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	888	74.9	226	11 Q9DCM2	Q9DCM2 mus musculus
2	198	16.7	195	16 Q9I714	Q9I714 pseudomonas
3	195.5	16.5	207	2 Q93NA9	Q93NA9 burkholderi
4	164.5	13.9	194	2 Q52782	Q52782 rhizobium 1
5	149	12.6	195	2 Q9X9Q7	Q9X9Q7 spingomona
6	147.5	12.4	206	16 Q9AAF8	Q9AAF8 caulobacter
7	145.5	12.3	312	5 Q22312	Q22312 caenorhabdi
8	140	11.8	197	2 Q85994	Q85994 spingomona
9	135	11.4	196	2 Q92HH4	Q92HH4 burkholderi
10	114	9.6	253	16 Q9AXH0	Q9AXH0 alcaligenes
11	114	9.6	253	16 Q9A923	Q9A923 caulobacter
12	113	9.5	241	2 Q93HB9	Q93HB9 streptomyces
13	106	8.9	199	2 Q923X5	Q923X5 ralstonia s
14	105	8.9	196	2 Q93CP1	Q93CP1 burkholderi
15	97.5	8.2	238	2 Q92167	Q92167 pseudomonas
16	91	7.7	581	16 Q9WXQ0	Q9WXQ0 thermotoga

17	90.5	7.6	410	17 Q26322	Q26322 methanother
18	90.5	7.6	600	16 Q92W85	Q92W85 rhizobium m
19	89	7.5	199	2 Q51499	Q51499 pseudomonas
20	87	7.3	926	16 Q911V3	Q911V3 pseudomonas
21	85	7.2	323	2 Q93H27	Q93H27 streptomyces
22	83.5	7.0	308	2 Q92NM7	Q92NM7 exiguobacte
23	83.5	7.0	1007	10 Q92VD4	Q92VD4 arabisidopsis
24	83	7.0	863	12 Q93124	Q93124 human calic
25	82.5	7.0	263	2 Q87197	Q87197 thermus the
26	82	6.9	203	2 Q07298	Q07298 pseudomonas
27	82	6.9	296	2 Q52799	Q52799 rhizobium 1
28	81.5	6.9	206	2 Q93CM5	Q93CM5 pseudomonas
29	80.5	6.8	413	10 Q9LUP2	Q9LUP2 arabisidopsis
30	80	6.8	221	5 Q09355	Q09355 nosome locu
31	80	6.8	402	17 Q26282	Q26282 methanother
32	80	6.8	691	5 Q95R88	Q95R88 drosophila
33	80	6.8	704	5 Q9VVT8	Q9VVT8 drosophila
34	80	6.8	821	2 Q30926	Q30926 escherichia
35	79	6.7	230	16 Q9RSB4	Q9RSB4 delnococcus
36	78.5	6.6	223	17 Q973S1	Q973S1 sulfolobus
37	78.5	6.6	333	16 Q9HYF7	Q9HYF7 pseudomonas
38	78.5	6.6	493	10 Q9SVJ3	Q9SVJ3 arabisidopsis
39	78.5	6.6	693	5 Q9VS91	Q9VS91 drosophila
40	78.5	6.6	778	16 Q92VY3	Q92VY3 rhizobium m
41	78.5	6.6	1072	6 Q77690	Q77690 bos taurus
42	78	6.6	248	17 Q29435	Q29435 archaeoglob
43	78	6.6	313	17 Q9YF18	Q9YF18 aeropyrum p
44	78	6.6	338	16 Q84200	Q84200 chlamydia t
45	78	6.6	687	11 Q9EPX1	Q9EPX1 mus musculu

ALIGNMENTS

RESULT 1

Q9DCM2 PRELIMINARY; PRT; 226 AA.

AC Q9DCM2; Q9DCM2; 01-JUN-2001 (TREMREL. 17, Created)

DT 01-JUN-2001 (TREMREL. 17, Last sequence update)

DT 01-JUN-2001 (TREMREL. 17, Last annotation update)

DE 0610025119RIK PROTEIN.

GN 0610025119RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=KIDNEY;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y., et al.

Functional annotation of a full-length mouse cDNA collection."

Nature 409:685-690(2001).

EMBL: AK002661; BAB22268.1; -

MGI:1923513; 0610025119RIK.

```

SQ SEQUENCE 226 AA; 25704 MW; 480332FD6D18ABCD CRC64;

Query Match 74.9%; Score 888; DB 11; Length 226;
Best Local Similarity 71.2%; Pred. No. 1.7e-75;
Matches 161; Conservative 33; Mismatches 32; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFYDVLSPYSWLGFEILCRQYNTWNLQRLPSLITGIMKDSGNKPPGLP 60
DB 1 MGPAIRLELFYDVLSPYSWLGFEVLCRYQHLWNKQLQRPRTLIAGIMKDSGNOPPAWVP 60
QY 61 RKGLYANDLKLRLHHLQIPITHFPRDLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120
DB 61 RKQGYIFKEIPLKQFFQVPLNIPKDFETGVKGSINAMRFLTAVNLEHPEMLEKASRE 120
QY 121 LWMRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLETTAAACRYGA 180
DB 121 IWMRVSRNEDITEYQISILAAAVKAGMSTAQAQHFLEKISTQOVKNKLETTAAACRYGA 180
QY 181 FGLPTTVAHVGDQTHMLFGSDRMELLALHGEKWMGPPIPPAVNARL 226
DB 181 FGLPTTVAHVGDGTYMLFGSDRLLELLAYLLGEKWMGPVPTANARL 226

RESULT 2
Q91714 PRELIMINARY; PRT; 195 AA.
AC Q91714;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HYPOTHETICAL PROTEIN PA0118.
GN Pseudomonas aeruginosa.
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AF004450; AAC03508.1;
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 21473 MW; CA3C029592EEB41D CRC64;

Query Match 16.7%; Score 198; DB 16; Length 195;
Best Local Similarity 26.1%; Pred. No. 9.7e-11;
Matches 55; Conservative 41; Mismatches 93; Indels 22; Gaps 5;

QY 4 LPTVELFYDVLSPYSWLGFEILCRQYNTWNLQRLPSLITGIMKDSGNKPPGLP 63
DB 1 MSKQIEFFDFGSPPTVLAQTLPRIAAHAGASIAWRPMLLGGVFKATGNHSPLEVPAKG 60
QY 64 LYMANDLKLRLHHLQIPITHFPRDLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASR 119
DB 61 RYTLHLARVAKRYGVPLAFNPAFPINTLTLM-----RGAQGYLGG-----EGFQYLYK 109
QY 120 LWMRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLETTAAACRYG 179
DB 120 ELWVRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLETTAAACRYG 179
DB 110 AVEALWVRQNLGKPEVAQVLAEGFDPDE---FLVLVGDEQVKEGLKATTEAAVRG 166
```

```

QY 180 AFGLPTTVAHVGDQTHMLFGSDRMELLALHLL 210
DB 167 VFGAPSEFV---GDQLEFFGQDRLDFAEVL 193

RESULT 3
Q93NA9 PRELIMINARY; PRT; 207 AA.
AC Q93NA9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE DBTD.
GN Burkholderia sp. DBT1.
OS Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=161152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBT1;
RA Di Gregorio S., Zocca C., Vallini G.;
RT "Burkholderia sp. strain DBT1 isomerase DbtD, dioxygenase (ISP) alpha
RT subunit DbtAC, dioxygenase (ISP) beta subunit DbtAD, dihydrodiol
RT dehydrogenase DbtB, putative NADH:FMN oxidoreductase and
RT transcriptional regulator DbtR."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF380367; AAK62352.1;
SQ SEQUENCE 207 AA; 24059 MW; 1E51731CBEACDD58 CRC64;

Query Match 16.5%; Score 195.5; DB 2; Length 207;
Best Local Similarity 24.3%; Pred. No. 1.8e-10;
Matches 52; Conservative 49; Mismatches 94; Indels 19; Gaps 6;

QY 4 LPTVELFYDVLSPYSWLGFEILCRQYNTWNLQRLPSLITGIMKDSGNKPPGL--LPR 61
DB 1 MERNLEFFDFMSPFAYLAHOKLPGLAEQDFLEITYRPVELKQLKLAAGNVSPGNRDIPL 60
QY 62 KGLYMANDLKLRLHHLQIPITHFPRDLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121
DB 61 KLTLYLKDMERWAERYGVPIPPSSLESLSLNGAYYAI-----DKGKIQQYVSEV 111
QY 122 WMRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLETTAAACRYGAF 181
DB 112 WSETWNGRQLDDQAVVLDVLDKFGW---DEQDFIKFTASDEAQDRYDDGTQYAHHRGVF 168
QY 182 GLPTTVAHVGDQTHMLFGSDRMELL-AHLIGERK 214
DB 169 GVP-TIAI---CAEMWGNDRLEILKEHLRPKOW 198

RESULT 4
Q52782 PRELIMINARY; PRT; 194 AA.
AC Q52782;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE PUTATIVE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE, SIMILAR TO NAHD
DE AND DOXJ PRODUCTS ENCODED BY GENBANK ACCESSION NUMBERS U09057 AND
DE M60405.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UPM791;
RX MEDLINE=96328256; PubMed=8765742;
RA Brito B., Palacios J., Imperial J., Ruiz-Argueso T.;
RT "Identification of a gene for a chemoreceptor of the methyl-accepting
RT type in the symbiotic plasmid of Rhizobium leguminosarum bv. viciae
```

```

Db 162 AYQGVGAP--MNEVDDQ--IFWGNDRDLFESYL 193
RESULT 6
Q9AAAF8 PRELIMINARY; PRT; 206 AA.
AC Q9AAAF8;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0643.
GN CC0643.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter
OC Caulobacter
OX NCBI_TaxID=69394;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 19089 / CB15;
RC MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA DeBoy R.T., Dodson R.J., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Haft D.H.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005740; AAK22628.1; -
DR TIGR; CC0643; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 206 AA; 22058 MW; 634328E44B88489A CRC64;

Query Match 12.4%; Score 147.5; DB 16; Length 206;
Best Local Similarity 24.0%; Pred. No. 5.8e-06;
Matches 50; Conservative 43; Mismatches 86; Indels 29; Gaps 8;

Qy 6 RTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPPG---GLLP 60
Db 8 KSIDIFDGSFNAYLSWKLPEAARAGAAVRLPCLLGGIFKATGQAPMAFGVKG 67
Qy 61 RKGLYMANDLKLRHH---LQIPHPKDFLSVWLEKSLSAMRFLTAVNLEHPMLEK 116
Db 68 KMDYEMLETRFRFAAHLTAFTAFENPFPVN---TLLMRGMIAAQIGVA-----E 115
Qy 117 ASRELWNR-VWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNQLKETEA 175
Db 116 AYLEAMKGMWEDGLKDDPEVFVATANAAGL---DGAALLAATGDAEVKAEVLVANTEA 172
Qy 176 CRYGAFGLPITVAHVGDQTHMLFGSDRM 203
Db 173 VARGTEGIPTFV---GE-EIFFGKERL 196

RESULT 7
Q22312 PRELIMINARY; PRT; 312 AA.
ID Q22312;
AC Q22312;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE T07E3.3 PROTEIN.
GN T07E3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC Wilton R., Alnscough R., Anderson K., Baynes C., Berks M.,

```

```

RT UPW791.;
PL Blochim. Biophys. Acta 1308:7-11(1996).
DR EMBL; U23040; AAC44312.1; -
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
KW Isomerase.
SQ SEQUENCE 194 AA; 21153 MW; 0421C131B37BA435 CRC64;

Query Match 13.9%; Score 164.5; DB 2; Length 194;
Best Local Similarity 25.2%; Pred. No. 1.4e-07;
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;

Qy 6 RTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64
Db 2 RILDFYDYSRPSYFLALSQVRK---MDVEIAFHPLEIGDLMKQGVNPTSITCAPKR 57
Qy 65 YNANDLKLRLHQLIPIHPKDFLSV-----MLEKGSLSAMRFLTAVNLEHPMLEK 116
Db 58 YVMTDIQWAVHYGVSLNHPOLLEIDASRLRLATLVAGQGLAMP--TAV----- 105
Qy 117 ASRELWNRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNQLKETEAAC 176
Db 106 --PAFNAHNSAPATAAEVAVLGAGLDAEE---LAERMDPEAAQDLDEATANAV 160
Qy 177 RYGAFLPITVAHVGDQTHMLFGSDRMELL-AHL 209
Db 161 SRGVFGAPTFLV-----GDEMFFGNDRLHFMQGH 190

RESULT 5
Q9X9Q7 PRELIMINARY; PRT; 195 AA.
ID Q9X9Q7;
AC Q9X9Q7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.
GN NSAD.
OS Sphingomonas sp.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=28214;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=BN6;
RC Keck A.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; U65001; AAD45416.1; -
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
KW Isomerase.
SQ SEQUENCE 195 AA; 22064 MW; 7E6FBAC96B3C6316 CRC64;

Query Match 12.6%; Score 149; DB 2; Length 195;
Best Local Similarity 23.1%; Pred. No. 3.9e-06;
Matches 50; Conservative 43; Mismatches 91; Indels 32; Gaps 7;

Qy 4 LPRTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPPG---LLP 60
Db 1 MTKTIDFYDFISPSYLAQVKLPDLARRTGCVEYRPIDPEAKIAGNYSNREYVP 60
Qy 61 RKGLYMANDLKLRHLQIPIHPKDF-----LSVMLEKGSLSAMRFLTAVNLEHPML 114
Db 61 KIKVMA-DLERWAAYEVLTPFASPACSDNWCNCAALYARQDAEAVTAA----- 111
Qy 115 EKASRELWNRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNQLKETEA 174
Db 112 -----YIRIWIGIDIPRONELRGCAEDVGLDAD---ALCEFVRSAGGGEYRKARTO 161
Qy 175 ACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHL 210

```


RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "The C. elegans genome project: Contiguous nucleotide sequence of over
 RT two megabases from chromosome III.",
 RL Nature 0:0-0(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Favello A.;
 RT "The sequence of C. elegans cosmid T07E3";
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U13643; AAA21082.1;
 SQ SEQUENCE 312 AA; 35436 MW; 0FB8561F57687120 CRC64;

Query Match 12.3%; Score 145.5; DB 5; Length 312;
 Best Local Similarity 22.6%; Pred. No. 1.6e-05;
 Matches 51; Conservative 46; Mismatches 100; Indels 29; Gaps 6;
 QY 1 MGPLPRTVELFVDVLSVPSW-----LGFELCRVQ-----NINWNLQ 38
 DB 1 MAPLR-VKCYFDVCPNSWITTAQTSNLSLFRIDFEPVCFKIGILHNAQINRRQ 59
 QY 39 LRPSLI--TGIMKDSGNKPPGLPRKGLYMAND---LKLRLHLLQIPHFPKDLSVME 93
 DB 60 VHSRLMTKIEVPEQAESEETLSENGILQKIDRGKLCICERVPVVDKNTYKTA 119
 QY 94 KGSLSAMRFLTAVNLEHPEMLEKASRELWVRVWSRNETITEPQSILAAAEKAGMSAEQAO 153
 DB 120 RGSVIPQLFSLRSQYDPOLYEKAHHLGKRLVQRLPVHYGCHMSTVCRELGISFKIAE 179
 QY 154 GLLEKPIATPKYNOLKTEAEACRYGAFGLPITVAHVD--GQTHMLF 198
 DB 180 DIVARLSSPENKSLHKNCKEAVDFKLTEAPGLILLTDEGOTIKIF 225

RESULT 8
 O85994
 ID- O85994 PRELIMINARY; PRT; 197 AA.
 AC O85994;

DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE- 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.
 GN NAHD.
 OS Sphingomonas aromaticivorans.
 OG Plasmid pNL1.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Novosphingobium.
 OX NCBI_TaxID=48935;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=F199;
 RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,
 RA Sensen C.W., Gaasterland T., Saffer J.D., Fredrickson J.K.;
 RT "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas
 RT aromaticivorans strain F199";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF079317; AAD04010.1;
 DR InterPro; IPR004287; HCCA_isomerase.
 DR Pfam; PF03046; HCCA_isomerase; 1.

KW Isomerase; Plasmid.
 SQ SEQUENCE 197 AA; 21998 MW; C0CD74B7106C07CE CRC64;
 Query Match 11.8%; Score 140; DB 2; Length 197;
 Best Local Similarity 23.1%; Pred. No. 2.8e-05;
 Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;
 QY 4 LPRTVELFYDLSVPSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPG---LLP 60
 DB 1 MTRTIDFYDFISFVSVAQLKLPFIARAAGCTVDYWPIDIPAKTAAGYGPSNRVLP 60
 QY 61 RKGLYMANDLKLRLHLLQIPHFPKDF-----LSVMELEKGSLSAMRFLTAVNLEHPEML 114
 DB 61 -KIKVMKADLERWAERYGVPLTFPASFACADWNCVLFAREHGKAEAFVT----- 109
 QY 115 ERASRELWVRVWSRNETITEPQSILAAAEKAGMSAEQAOGLLEKIATPKYNOLKTEAE 174
 DB 110 -----DAYRIWGGIDPGDRNELAACIAAGL---DPAALIAFVESPAGQNEYRKARSQ 161
 QY 175 ACRYGAFGLPITVAHVDGQTHMLFGSDRMELLALL 210
 DB 162 AIQRGVYGAPL--MFVDDQ--IFWGNDRDLFLAEVL 193

RESULT 9
 Q92HH4

ID- Q92HH4 PRELIMINARY; PRT; 196 AA.
 AC Q92HH4;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE ISOMERASE PHND.
 GN PHND.
 OS Burkholderia sp. RP007.
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 OX NCBI_TaxID=83784;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=RP007.
 RX MEDLINE=99102216; PubMed=9882667;
 RA Laurie A.D., Lloyd-Jones G.;
 RT "The phn genes of Burkholderia sp. Strain RP007 constitute a divergent
 RT gene cluster for polycyclic aromatic hydrocarbon catabolism";
 RL J. Bacteriol. 181:531-540(1999).
 DR EMBL; AF061751; AAD09871.1;
 DR InterPro; IPR004287; HCCA_isomerase.
 DR Pfam; PF03046; HCCA_isomerase; 1.
 SQ SEQUENCE 196 AA; 21908 MW; 58716E9039BEA76B CRC64;

Query Match 11.4%; Score 135; DB 2; Length 196;
 Best Local Similarity 24.0%; Pred. No. 8.1e-05;
 Matches 50; Conservative 35; Mismatches 103; Indels 20; Gaps 5;

QY 7 TVELFYDLSVPSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGL--LPRKGL 64
 DB 2 TIDFFDFLSPYAVLARHLRTQVAALHCAIYKPIDLARAKLAIGNTGFANRDPVKLA 61
 QY 65 YMANDLKLRLHLLQIPHFPKDFLSVMELEKGSLSAMRFLTAVNLEHPEMLEKASRELWVR 124
 DB 62 YVVEDLKRWAARYRIPIEFIEFKNFNTKRNVTG-----FYAARGOQADYVQAVHLWAGE 116
 QY 125 VWSRNETITEPQSILAAAEKAGMS-AEQAOGLLEKIATPKYNOLKTEAEACRYGAFGL 183
 DB 117 GGAPDDD-----AALRSYAVSGWGAADFLRFLDSSEATAYNESTLEASIGVFGV 168
 QY 184 PITVAHVDGQTHMLFGSDRMELLALL 211
 DB 169 PTMAVGGRD-----MWWGNDRIDFLFLEHLG 192

```

RESULT 10
99WXHO PRELIMINARY; PRT; 196 AA.
ID Q9WXHO
AC Q9WXH0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
ISOMERASE.
PHND.
GEN PHND.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Alcaligenes.
OC NCBI_TaxID=511;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-AFK2;
RC Kiyohara H., Tabata Y., Takizawa N.;
RA "A phenanthrene degradative gene cluster in Alcaligenes faecalis
RT AFK2."
RT Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RL EMBL; AB024945; BAA76325.1; -.
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
DR SOURCE NCBI; 196 AA; 22355 MW; C02759783BEE2E2 CRC64;

```

Query Match	9.6%	Score 114;	DB 2;	Length 196;
Best Local Similarity	21.1%	Pred. No. 0.0077;		
Matches	44;	Conservative	40;	Mismatches 99; Indels 26; Gaps 7;
QY	8	VELFYDLSPSYSLWGF----	EILCRYQNIWINLQLRPSLITGIMKDSGNKPPGL--	LPR 61
Db	2	LSFYDFVSFYSYLASIRLPEIVQRY----	GISVSYKPIDIACAKRAIGNVGPNSRDMPV	57
QY	62	KGLYMANDLLKRLHIHLOIPTHPKDFLSVMLEKGLSAMRFLTAVNLEHPEMLEKASREL	121	
Db	58	KLTHLSRDLQWQAQRYGTPKLKPPSPFSRRLNTGF----	FYAAGEAREAEYVRRR-----	108
QY	122	WMRYWSRNEDITEPQSTILAAEKAGMSAQOGLLEKIATPKVNQKLUKETTEAACRYCAF	181	
Db	109	FHLTWGMQAFSGEPFVLRKSIASEMGWNVDDFMQFTDSV----	DGANEYKQSIDEGIARSVF	165
QY	182	GLPTTVARVDGQTHMLFGSDRMELLAHL	210	
Db	166	GVPWVVI-----	GDEMWMGNDRLDFVDEYL	190
RESULT	11			
Q9A923		PRELIMINARY:	PRT:	253 AA.
ID	Q9A923			

DI 01-JUN-2001 (TRENBLERel. 19, Last annotation update)
DT 01-DEC-2001 (TRENBLERel. 19, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE, PUTATIVE.
GN CC1171.
OS Caulobacter crescentium.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin M.B., Khouri H., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Kohnen J., Berry K.,
RA Utterback T., Tran K., Wolf A., Yamamoto J., Fraser C.M.,
RA Salzberg S.L., Venter J.C., Shapiro L., Ermolaeva M., White O.,
RT "Complete genome sequence of *Caulobacter crescentium*."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

QY 217 -PIPPA 221
|:|
Db 225 LPVPA 230

RESULT 13

Q923X5 ID Q923X5 PRELIMINARY; PRT; 199 AA.
AC Q923X5;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2-HYDROXYCHROMENE CARBOXYLATE ISOMERASE.
GN NAGD.
OS Ralstonia sp. U2.
OG Plasmid pMWU2.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=70356;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U2;
RX MEDLINE=98233751; PubMed=9573207;
RA Fuenmayor S.L., Wild M., Boyes A.L., Williams P.A.;
RT "A gene cluster encoding steps in conversion of naphthalene to
genitiste in Pseudomonas sp. strain U2";
RL J. Bacteriol. 180:2522-2530(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=U2;
RX MEDLINE=20576173; PubMed=11133965;
RA Zhou N.Y., Fuenmayor S.L., Williams P.A.;
RT "nag genes of ralstonia (Formerly pseudomonas) sp. Strain U2 encoding
enzymes for gentisate catabolism";
RL J. Bacteriol. 183:700-708(2001).
DR EMBL; AF036940; AAD12617.1; -
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
KW Isomerase; Plasmid.
SQ SEQUENCE 199 AA; 23067 MW; 219466E06AC04E47 CRC64;

Query Match 8.9%; Score 106; DB 2; Length 199;
Best Local Similarity 24.9%; Pred. No. 0.044;
Matches 52; Conservative 32; Mismatches 91; Indels 34; Gaps 8;

QY 8 VELFYDVLSPYSWLGFEILCRQNTQNIWNINLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65
Db 5 VDFYDFLSFSPSYLANHRLSVLAGRYGFSIQYHAIDLAKATAIGNIGPSNRDLKVLKY 64
QY 66 MANDKLLRHHLQIPIHFPHKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASR----- 119
Db 65 LKVDLQWADLYRPLVFPNPNRSRVNAG-----LYYPAAREFAEYVRLVF 112
QY 120 -ELMRVSRNEDITEPOSILA-AAEKAGMSAEQAQGLLEKIATPKVKNLKETEACR 177
Db 113 DSAWGKGWALDAD-----SLLAEVCDKLNWDLGERFDELSENAAKAYD---ETQQAID 164
QY 178 YGAFGLPITVAHVGDQTHMLFGSDRMELL 206
Db 165 RKVEGVP-TVFWDD---QMWGNDRLMFL 189

RESULT 14

Q93CPI ID Q93CPI PRELIMINARY; PRT; 196 AA.
AC Q93CPI;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE ISOMERASE.
OS Burkholderia sp. DBT1.

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=161152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBT1;
RA Di Gregorio S., Zocca C., Vallini G.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF404408; AAK96187.1; -
KW Isomerase.
SQ SEQUENCE 196 AA; 22515 MW; CB5B111CEA6C08E1 CRC64;

Query Match 8.9%; Score 105; DB 2; Length 196;
Best Local Similarity 20.5%; Pred. No. 0.054;
Matches 42; Conservative 40; Mismatches 105; Indels 18; Gaps 5;

QY 4 LPTVELFYDVLSPYSWLGFEILCRQNTQNIWNINLQRLPSLITGIMKDSGNKPPG--LLPR 61
Db 1 MTKNLFYDFISFSPSYLAHIKLPALARQYGYTISYHPIDIPTAKIAAGNPGPSNREVQS 60
QY 62 KGLYMANDKLLRHHLQIPIHFPHKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASREL 121
Db 61 KMKVLLADMRWAARYDVPLTFPGKFCGERWNIGVLYA-----DKKRDTEAYVTET 111
QY 122 WMRVSRNEDITEPOSILA-AAEKAGMSAEQAQGLLEKIATPKVKNLKETEACRYGAF 181
Db 112 YHRWGLGINPSDEGOLGEVAEKGWNAIE---FLTFVSSPEGQATFRKSCVEAHARGVF 168
QY 182 GLPITVAHVGDQTHMLFGSDRMELL 206
Db 169 GAPIMV---GE-EVWGNDRMLFL 189

RESULT 15

Q92167 ID Q92167 PRELIMINARY; PRT; 238 AA.
AC Q92167;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE.
GN NAHD.
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AN10;
RX MEDLINE=99365311; PubMed=10433976;
RA Bosch R., Garcia-Valdes E., Moore E.R.B.;
RT "Genetic characterization and evolutionary implications of a
chromosomally encoded naphthalene-degradation upper pathway from
Pseudomonas stutzeri AN10";
RL Gene 236:149-157(1999).
DR EMBL; AF039533; AAD02142.1; -
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
SQ SEQUENCE 238 AA; 26359 MW; 29B972C1CCF7C71B CRC64;

Query Match 8.2%; Score 97.5; DB 2; Length 238;
Best Local Similarity 23.5%; Pred. No. 0.35;
Matches 53; Conservative 37; Mismatches 103; Indels 33; Gaps 9;

QY 8 VELFYDVLSPYSWLGFEILCRQNTQNIWNINLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65
Db 19 VDFYDFLSFSPSYLANHRLSVLAGRYGFSIQYHAIDLAKATAIGNIGPSNRDLKVLKY 78
QY 66 MANDKLLRHHLQIPIHFPHKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRELWMR 124
Db 79 LMVDLKRWAELGLPELFPANYSORMNAGLYSAGETQTAAYV-----NTVFNA 128

